Development of a picosecond time-of-flight system in the ATLAS experiment

Hervé Grabas

Thèse de doctorat
de l'Université Paris Sud

spécialité
Particules, Noyaux, Cosmologie

présentée par

Hervé GRABAS

pour obtenir le grade de docteur
de l'Université Paris Sud

Développement d'un système de mesure de temps de vol picoseconde dans l'expérience ATLAS

Thèse soutenue le 3 Décembre 2013 devant le jury composé de:

Marco BRUSCHI (**rapporteur**)
Eric DELAGNES (**co-directeur de thèse**)
Bertrand GRANADO
Krzysztof PIOTRZKOWSKI
Patrick PUZO
Christophe ROYON (**directeur de thèse**)
Gary VARNER (**rapporteur**)

Thèse préparée au
Commissariat à l'énergie atomique
CEA de Saclay
DSM/Irfu, bât. 141
91 191 Gif-sur-Yvette CEDEX

Contents

Introduction

Since year 2009, when I first started working on picosecond timing at the University of Chicago, experimental physics has greatly benefited from industry development towards faster, cheaper, power-saving circuits. The precision of the time measurement, once limited by signal acquisition and processing of particles detector, has gained an order of magnitude in speed and in precision using ultra-fast analog memories, thus reaching the picosecond frontier.

Simultaneously, process improvements and cost reduction of fast photo-detectors brought by the spread of camera sensors in today's digital equipment, but also a long history of imaging device development (pixel detectors at CERN or medical scanners) and reliability improvement in the photo-multiplier industry, have made high precision, gain and resolution photo-detectors widely available to the physics community. Research and development such as the Large Area Picosecond Photo-Detector program are also working on improving the resolution and size of today's photo-detector, while keeping low production costs.

This combination of factors allows for new measurement possibilities in experimental physics.

At the speed of light, a precision of 1 picosecond in time-of-flight measurement corresponds to a precision of 0.3 mm in the traveled distance. This unprecedented precision in space allows for a whole new range of measurement in particle physics. In the actual exploratory context in particle physics: effort to observe unseen, but predicted, particles such as the Higgs boson (observed 4 July 2012) or discover new physic beyond the Standard Model, the number of simultaneously interacting particles in particle colliders has been considerably increased, reaching up to 100 simultaneous interactions at the LHC (compared to 6 at the Tevatron). This high number of pile-up events requires new identification methods for the multiple vertices created. One of which is time-of-flight, where the particles created are matched to their vertex of origin measuring their travel time from their vertex of origin to their detection point. A 1-picosecond resolution yielding to a 0.3 mm resolution on the vertex position, giving a sufficiently fine segmentation resolution compared with the 20 cm size of the bunch crossing.

In this thesis, we are studying, in the first Part, the installation and benefits of fast timing detectors for the ATLAS experiment at the LHC. We are looking at the possible improvements in sensitivity to beyond Standard Model anomalous couplings between photons and the W boson. In the second Part is presented the necessary improvements for production of fast photo-detectors and the signal processing required to extract precise timing.

In the third Part is presented the custom micro-electronics developed for precise time measurement (the SamPic ASIC) and the performances obtained with the ASIC.

Part I

High energy physics

1 The Large Hadron Collider

Contents

1.1 Introduction

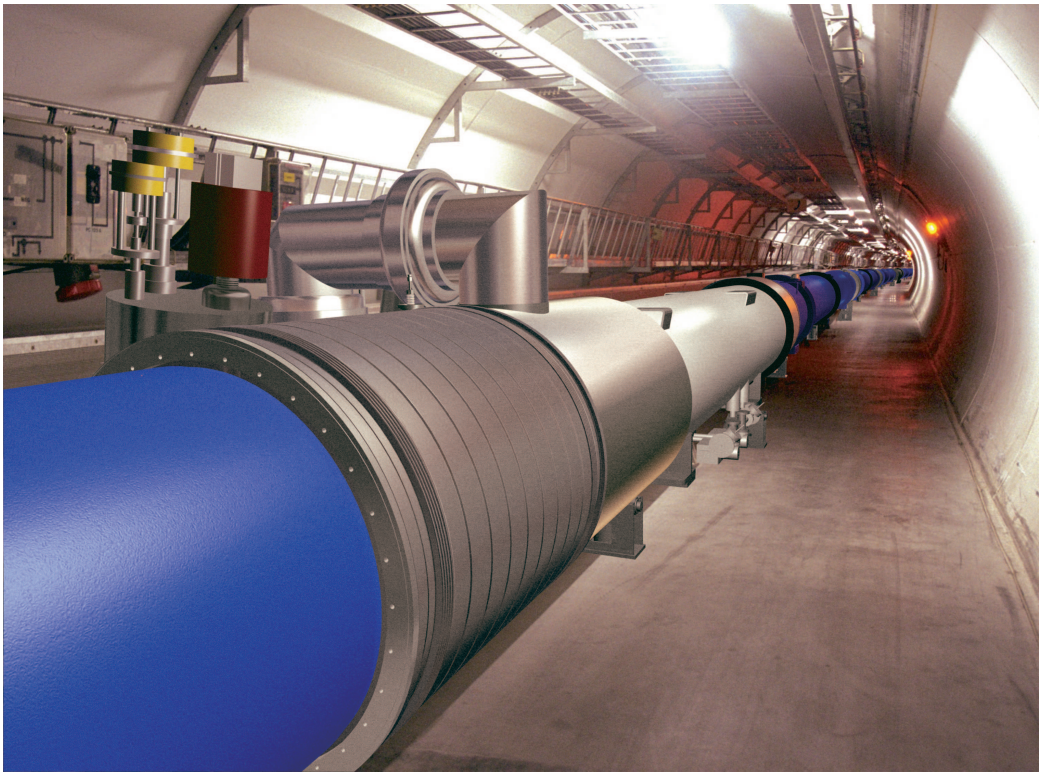


Figure 1.1 – Computer-generated image showing a section of the LHC tunnel after the magnets installation.

The Large Hadron Collider (LHC) [1, 2], installed in a tunnel at the border of

France and Switzerland, is the world's largest particle accelerator. It was started on September 10, 2008, after ten years of installation work. It has a circumference of 27 kilometers, divided in eight independent sectors. In the tunnel, 1232 superconducting dipole magnets are used to bend two high-energy particle beams. The beams travel in opposite directions in separate beam pipes kept at ultrahigh vacuum. The LHC principally collides protons but is also designed to work with heavy ions.

The LHC is not a perfect circle and is made of eight arcs and eight insertion regions. The arc are meant to bend the beam and contain 154 dipole magnets each. The insertions are straight sections joining the arcs. The connection between them is made in a transition region, the dispersion suppression region, reducing the machine dispersion (offsets in momentum) at the insertions. The layout of the straight section depends on the specific use of the insertion: physics (beam collision within an experiment), injection, beam dumping, beam cleaning.

1.2 The CERN accelerator complex

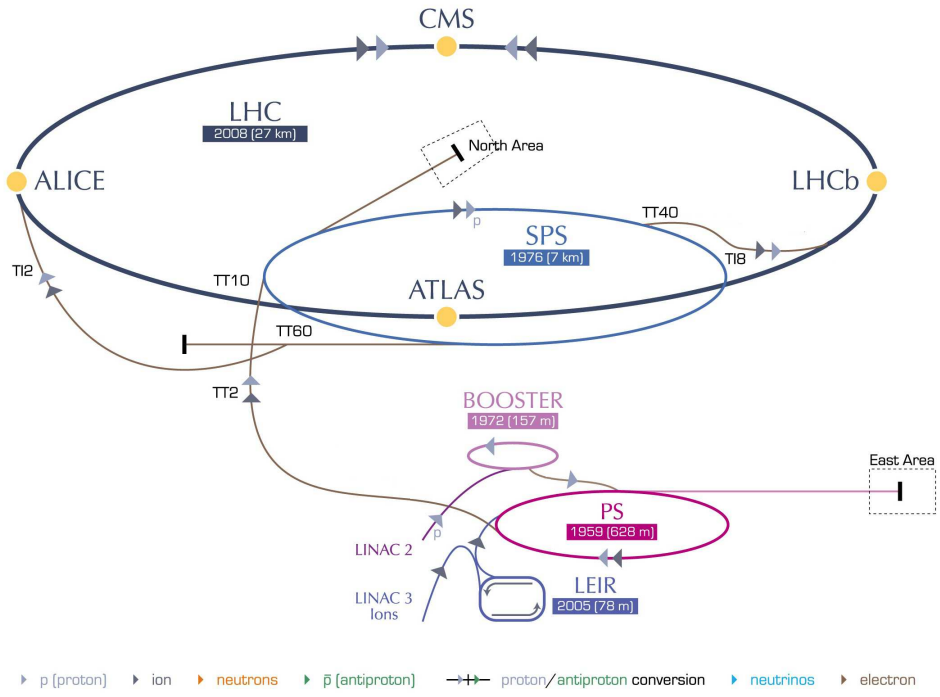


Figure 1.2 – LHC injection chain and main ring with the four collision points (yellow dots) where the main experiements are located (ALICE, ATLAS, CMS and LHCb) displayed.

The accelerator complex at CERN is a succession of machines increasing the energy of the particles, each machine injecting the beam into the next one. The LHC is the last element of the chain, each particle being accelerated up to the record energy of 4TeV (ultimately 7), making it the world's most powerful particle accelerator.

The proton acceleration process starts with hydrogen atoms taken from a simple bottle of hydrogen gas. An electric field is used to strip hydrogen atoms of their elec-

trons and the resulting ions are injected into RF cavities where they are accelerated to 750 keV. The particles are then accelerated by Linac 2, the first accelerator in the chain, to an energy of 50 MeV, before being sent to the Proton Synchrotron Booster (PSB), which accelerates the protons to 1.4 GeV, followed by the Proton Synchrotron (PS), which pushes the beam to 25 GeV. Protons are then sent to the Super Proton Synchrotron (SPS) where they are accelerated to 450 GeV. At this level of energy the beam is ready to be injected into the LHC. The LHC injection chain is shown in Figure 1.2.

The protons are transferred to the two beam pipes of the LHC via two transfer lines, the beam in one pipe circulating clockwise while the beam in the other pipe anticlockwise. It takes 4 minutes and 20 seconds to fill each LHC ring, and 20 minutes for the protons to reach their maximum energy of 4 TeV. Beams circulate for many hours inside the LHC beam pipes under normal operating conditions. The two beams are brought into collision inside four detectors - ALICE [3], ATLAS [4], CMS [5] and LHCb [6] - where the total energy at the collision point is presently equal to 8 TeV (14 TeV designed).

Besides the four main experiments, there are three smaller ones: TOTEM [7], LHCf [8] and MoEDAL [9]. Since there are only four beam intersection regions, the small experiments need to share them with the main ones. The TOTEM (TOTal Elastic and diffractive cross section Measurement) detectors share Point 5 with the CMS experiment and provide a coverage of the very forward rapidity region. The LHCf (LHC forward) shares Point 1 with the ATLAS experiment and is located 140 meters from the interaction region on both sides. Finally, the MoEDAL (Monopole and Exotics Detector At the LHC) experiment shares Point 8 with LHCb.

The beam is structured into 3564 bunches, each of them containing $\approx 10^{11}$ protons. The time spacing between the bunches is fixed at the Proton Synchrotron stage to 50 ns up to now (25 ns probably in 2015). The proton bunch length is 7.55 cm at the location of ATLAS and CMS and 16.7 μm in the transverse direction.

1.3 Number of interactions

At the location of the ATLAS and CMS experiments (interaction points one and five) the two particle beams trajectories are set on interception courses in a single beampipe and each proton bunch compressed as much as possible to increase the probability of collision.

We define the instantaneous luminosity of a particle accelerator $\mathcal{L}$ as the ration of the number of interaction per second in the detectors (or interaction rate $\frac{dR}{dt}$) to the total probability of interactions σ_{tot} (or total cross-section) [1].

$$\frac{dR}{dt} = \mathcal{L} \times \sigma_{tot} \tag{1.1}$$

And the luminosity $\mathcal{L}$, given the parameters of the beam can be calculated [10] as:

[1] Actually we only consider σ_{inel} for σ_{tot} since the elastic interactions are undetected

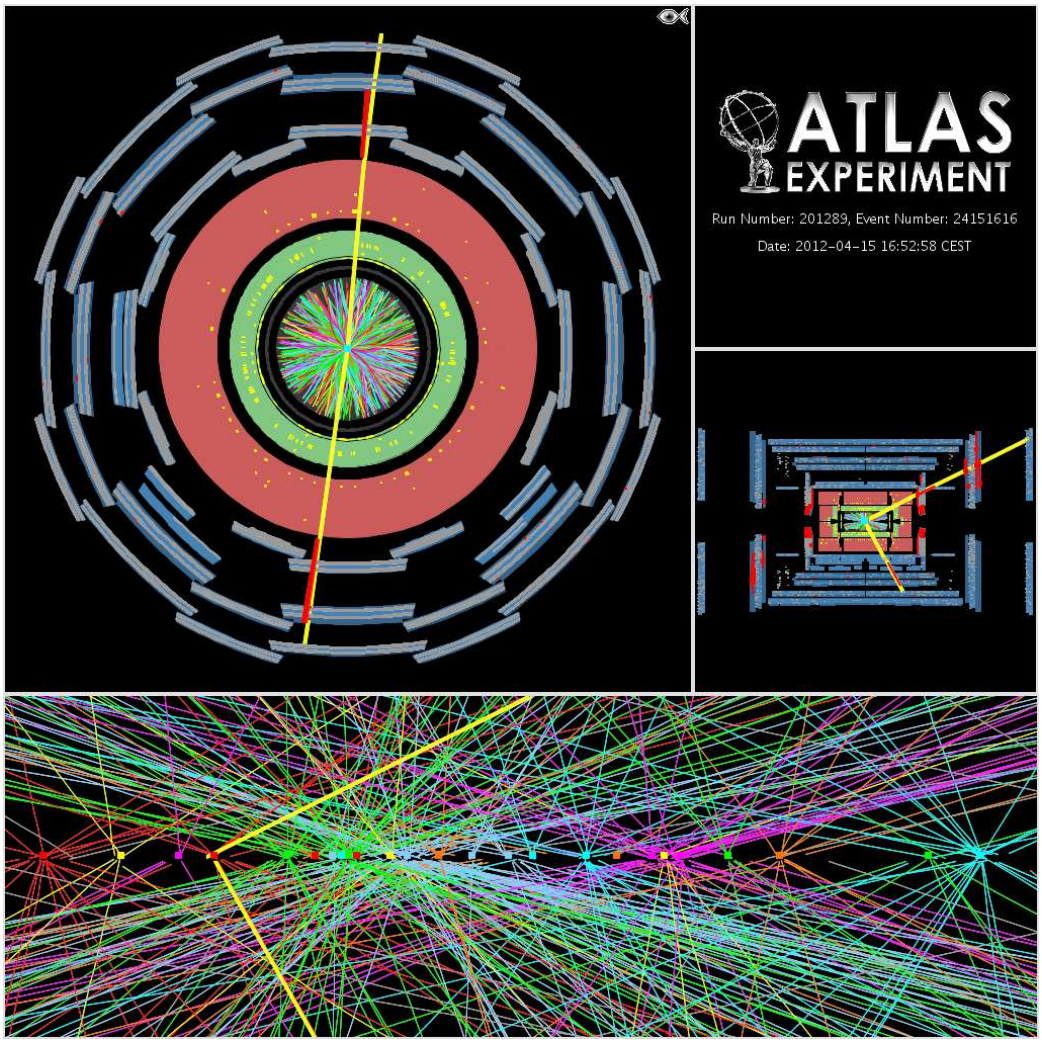


Figure 1.3 – Plot of high pile up event with 25 reconstructed vertices and a high p_T diboson event.

$$\mathcal{L} = \frac{kf}{4\pi\sigma_x\sigma_y}N^2 \tag{1.2}$$

where σ_x and σ_y are the beam transverse horizontal and vertical profile respectively, k the number of bunches per beam and f the revolution frequency.

Due to the continuous collisions the number N of particle decreases during a LHC run. The decay rate of the number of particles $\frac{dN(t)}{dt}$ is therefore negative and equal to the inverse of the interaction rate $\frac{dR}{dt}$.

$$\frac{dN(t)}{dt} = -\frac{dR}{dt} \tag{1.3}$$

Consequently, the luminosity decreases during the run and if N_0 is the initial number of particles and $\mathcal{L}_0$ the initial luminosity after injection, we have:

$$\mathcal{L} = \frac{\mathcal{L}_0}{(1 + {}^{t}/_{\tau})^2} \tag{1.4}$$

with τ the initial decay time of the beam intensity:

$$\tau = \frac{N_0}{k.\mathcal{L}_0.\sigma_{tot}} \tag{1.5}$$

If f is the average collision frequency, the mean number of interaction per bunch crossing μ is given by:

$$\mu = \frac{\sigma_{tot} \times \mathcal{L}}{f} \tag{1.6}$$

The average collision frequency f differs from the bunch-to-bunch spacing in time (25ns corresponding to $40MHz$), because not all of the available beam buckets of the ring are filled. The average time between two collisions is therefore higher (31.7ns) and the mean collision frequency $f = 31.5MHz$. Due to the luminosity decay the mean number of interactions per bunch crossing μ decreases through time.

For an initial luminosity $\mathcal{L} = 10^{34}cm^{-2}s^{-2}$ and a total cross-section $\sigma_{tot} = 100mb = 10^{-25}cm^2$ we obtain $\mu \approx 32$ interactions per bunch crossing [11]. For higher scheduled luminosity, this number will increase to values between $\mu = 50$ to $\mu = 100$.

Bibliography

[1] Oliver Brüning, Paul Collier, Ph Lebrun, Stephen Myers, Ranko Ostojic, John Poole, and Paul Proudlock. *LHC design report(Volume I, The LHC main ring)*. *Reports- CERN*, 2004.

[2] Lyndon Evans and Philip Bryant. *LHC machine*. *Journal of Instrumentation*, 3(08):S08001, 2008.

[3] Kenneth Aamodt, A Abrahantes Quintana, R Achenbach, S Acounis, D Adamová, C Adler, M Aggarwal, F Agnese, G Aglieri Rinella, Z Ahammed, et al. *The ALICE experiment at the CERN LHC*. *Journal of Instrumentation*, 3(08):S08002, 2008.

[4] Georges Aad, E Abat, J Abdallah, AA Abdelalim, A Abdesselam, O Abdinov, BA Abi, M Abolins, H Abramowicz, E Acerbi, et al. *The ATLAS experiment at the CERN large hadron collider*. *Journal of Instrumentation*, 3(08):S08003, 2008.

[5] S Chatrchyan, G Hmayakyan, V Khachatryan, AM Sirunyan, W Adam, T Bauer, T Bergauer, H Bergauer, M Dragicevic, J Erö, et al. *The CMS experiment at the CERN LHC*. *Journal of Instrumentation*, 3(08):S08004, 2008.

[6] A Augusto Alves Jr, LM Andrade Filho, AF Barbosa, I Bediaga, G Cernicchiaro, G Guerrer, HP Lima Jr, AA Machado, J Magnin, F Marujo, et al. *The LHCb detector at the LHC*. *Journal of Instrumentation*, 3(08):S08005, 2008.

[7] Giovanni Anelli, G Antchev, P Aspell, V Avati, MG Bagliesi, V Berardi, M Berretti, V Boccone, U Bottigli, M Bozzo, et al. *The Totem experiment at the cern large hadron collider*. *Journal of Instrumentation*, 3(08):S08007, 2008.

[8] Oscar Adriani, L Bonechi, M Bongi, G Castellini, R D'Alessandro, DA Faus, K Fukui, M Grandi, M Haguenauer, Y Itow, et al. *The LHCf detector at the CERN Large Hadron Collider*. *Journal of Instrumentation*, 3(08):S08006, 2008.

[9] James Pinfold, K Kinoshita, R Soluk, J Jakubek, Y Yao, D Felea, DH Lacarrère, L Patrizii, S Cecchini, T Hott, et al. *Technical design report of the moedal experiment*. Technical report, 2009.

[10] Werner Herr. *Effect of missing head-on collision on beam-beam effects in the LHC*. Technical report, CERN-LHC-Project-Note-321, 2003.

[11] F Zimmermann. *Parameter Space Beyond 1034*. Technical Report EuCARD-CON-2010-041, Nov 2010.

2
The ATLAS detector

Contents

2.1 Introduction

The purpose of the four main experiments installed at the LHC - ATLAS, CMS, ALICE and LHCb - is to detect and identify the secondary particles produced in collisions at four interaction points in order to reconstruct and analyze the physics of each event. To do so, the detector goal is to measure the position, charge, speed, mass and energy of all particles created. Therefore, particle detectors needs to be hermetic, so that only a small fraction of the particles can escape undetected from the detector. For engineering convenience, detectors adopt the "barrel plus endcap" design, where a cylindrical detector covers the central region and two flat circular "endcaps" cover the angles close to the beam (the forward region).

In Section 2.2, we present the basic concepts of particle detection in high energy physics experiments. Section 2.3 describes the ATLAS detector in more detail.

2.2 Particle detection and identification

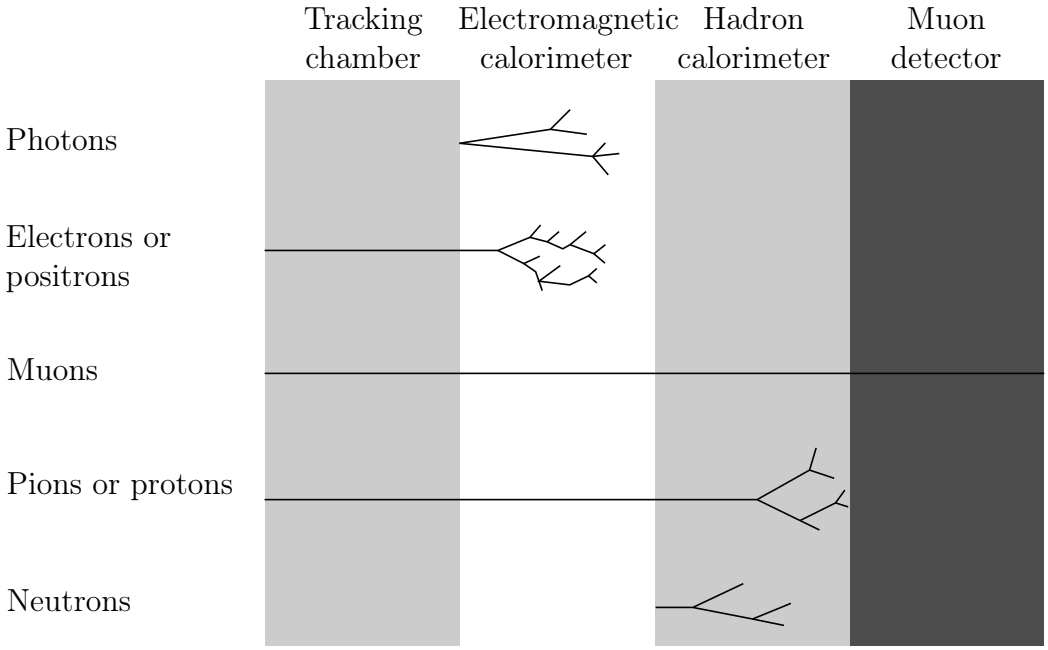


Figure 2.1 – Illustration of the different layers of a particle detector and the signature of the different kinds of particles [1].

A particle detector is essentially made of these types of detectors:

- Tracking detectors. Measuring the tracks that electrically charged particles are leaving ionizing the matter.
- Calorimeters. Measuring the energy (and position) of leptons and hadrons.
- Muons detectors, dedicated to measure muons that can travel undifferentiated through all previous detectors.

2.2.1 Tracks

Tracking the particle trajectory from their original creation point allows reconstruction of several properties of the particle produced:

- **Vertex** As we saw, at the LHC, there are, in average, several interactions occurring at each bunch crossing (depending on the luminosity and number of particles per bunch). Each interaction - or *event* - takes places at a precise, but random, position in bunch crossing (see Figure 1.3) and produces hundred's of particles. Therefore, reconstructing what happened during an *event* implies matching all the particles to their point of origin or vertex. The determination of all vertices is done by finding the intersection point of all produced particles.

- **Charge** By putting the particle detector inside a magnetic field the charged particles will have trajectories bent in one way for positively charged particle and the opposite for negative, thus allowing identification of charged particles.
- **Momentum** A charged particle with a high momentum will travel almost in straight line in the magnetic field, while low momentum particle will make tight spiral. By measuring the radius of the tracks curvature it is therefore possible to measure the particle momentum.

2.2.2 Calorimeters

Tracking detector cannot see neutral particles such as photons or neutrons and are inefficient to measure the energy of particles crossing them. Calorimeters are composed of high-density material that causes particle shower on impact. The measurement of the full shower shape, constitution and number of particles allows determination of the incident particle energy and its identification. There are two types of calorimeters: electromagnetic (ECAL) and hadronic (HCAL). They use different types of material, the ECAL generally absorbing completely electrons and photons interacting via the electromagnetic force (hence the name). Strongly interacting particle such as protons and pions, may begin to lose energy in the ECAL but will be fully stopped in the HCAL.

Calorimeters typically consists of layers of *absorbing* high density metal (lead for instance) interleaved with layers of *active* medium such as lead-glass or liquid argon used to measure the particle showers created in the absorbing material and measure them. There are also homogeneous calorimeters where the active layer is also the absorbing medium, for example lead glass calorimeters.

2.2.3 Tracker and calorimeter association

It is the association of calorimeters and trackers with the muon system that allow to completely identify particles as shown in Figure 2.1.

Photons and electrons, while having the same signature in the ECAL are differentiated using the tracking system, identifying the electron as a charged particle. A similar differentiation is done for the protons to neutrons identification, neutrons leaving no tracks.

2.2.4 Muon system

An additional dedicated tracker is added on the outer shell of the detector for the muon identification. Muons are able to travel undifferentiated trough dense material (ECAL and HCAL) and will be identified and measured in the muon spectrometer.

Since muons leave clean tracks in all the detector parts (see Figure 2.1), they often offer the cleanest reconstruction and measurement methods of primary particle, for example for the muon decay of the W boson:

$$W^{\pm} \rightarrow \mu^{\pm}\nu_{\mu}$$

2.3 The ATLAS detector

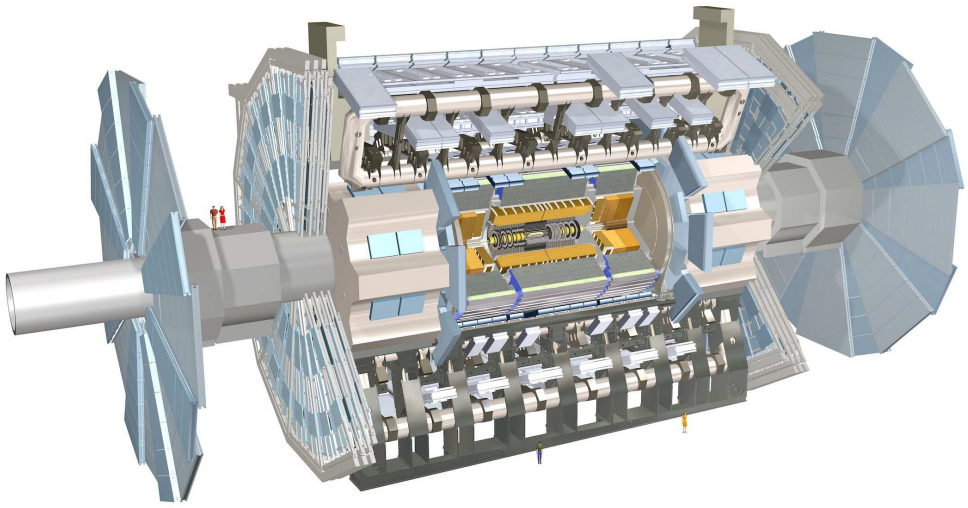


Figure 2.2 – General diagram of the ATLAS main detector, with a total length 22 m and a radius of 11 m.

A cut-away view of the ATLAS (A Toroidal LHC ApparatuS) detector is shown in Figure 2.2, revealing its internal structure. From the innermost part to the outer, it is composed of the inner tracking system, electronic and hadronic calorimeters and lastly the muon detectors.

The goal of the inner tracking detector is to identify charged particle vertices. It is surrounded by a thin superconducting solenoid generating a uniform 2 Tesla magnetic field. The calorimeter in ATLAS uses liquid argon as an active layer both for the electromagnetic and hadronic calorimeters. Notably the signal sampling is done using Sampling Capacitor Arrays (see SamPic architecture in Chapter 8). The ATLAS muon system is placed on both sides of an air-core toroid magnetic system generating 0.5 Tesla in its center. The ATLAS magnet system is shown in Figure 2.4.

Overall the ATLAS detector is a cylinder with a total length of 42 meters and a radius of 11 meters.

2.3.1 Coordinate system

The coordinate system of ATLAS is a right-handed coordinate system centered at the ATLAS interaction point. The x-axis is pointing towards the center of the LHC ring and the z-axis along the LHC tunnel. Due to the slope of the beam at the ATLAS

interaction points the y-axis is slightly tilted from the vertical at that point. We usually define the pseudo-rapidity of the particle from the primary vertex as:

$$\eta = -\ln \tan \frac{\theta}{2} \tag{2.1}$$

where θ is the polar angle of the particle direction measured from the positive z-axis. In the limit where the particle is travelling close to the speed of light, or in the approximation that the mass of the particle is nearly zero, pseudorapidity is numerically close to the particle rapidity y:

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z} \tag{2.2}$$

with E the particle energy and p_z the component of the momentum along the beam axis. The transverse momentum p_T, is defined as the momentum perpendicular to the LHC beam axis (or projection of the momentum on the xy plane).

2.3.2 The inner detector

This detector is made of three distinct parts [2, 3]:

- A Silicon pixel detector.
- A Semi-conductor tracker (SCT).
- A Transition Radiation Tracker (TRT).

The pixel detector and the SCT cover a region such as $|\eta| \leq 2.5$ whereas the TRT covers the region $|\eta| \leq 2.0$. The structure of this detector is shown in Figure 2.3.

The pixel detectors are made of layers of silicon tiled together and forming a barrel geometry around the interaction point. They each have a very high granularity and fine segmentation in order to accommodate with the high density of tracks. It is used to reconstruct very precisely the position of the primary and secondary vertexes. Composed of three layers placed at $R = 50.5$, 88.5 and 122.5 mm (R distance to the beam axis), the precision is 10 μm in $R\phi$ (ϕ the azimuthal angle) and 115 μm in z.

The second sub-detector is the Semi-Conductor Tracker (SCT), located around the pixel detector and made of micro-strips tilted from each other to reconstruct track hits. It consists of 4 layers in the barrel and 9 layers in the end-caps. The SCT provides a resolution of 17 μm in $R\phi$ and 580 μm in z.

The outermost detector of the inner part of ATLAS is the Transition Radiation Tracker (TRT). It is built of "straw" tubes filled with a xenon-based gas mix. A conductive coating deposited on the inner part the straw acts as a cathode, while a thin wire stretched inside the wire is the anode. Charged particles traversing the straw ionize the gas and the signal is collected on the anode. Additionally, transition radiation is emitted when the particle with a high Lorentz factor ($\gamma = E/m$) crosses materials with different electrical properties - an electron has a much higher probability to emit transition radiation than a pion. This difference in energy allows electrons to be distinguished from hadrons.

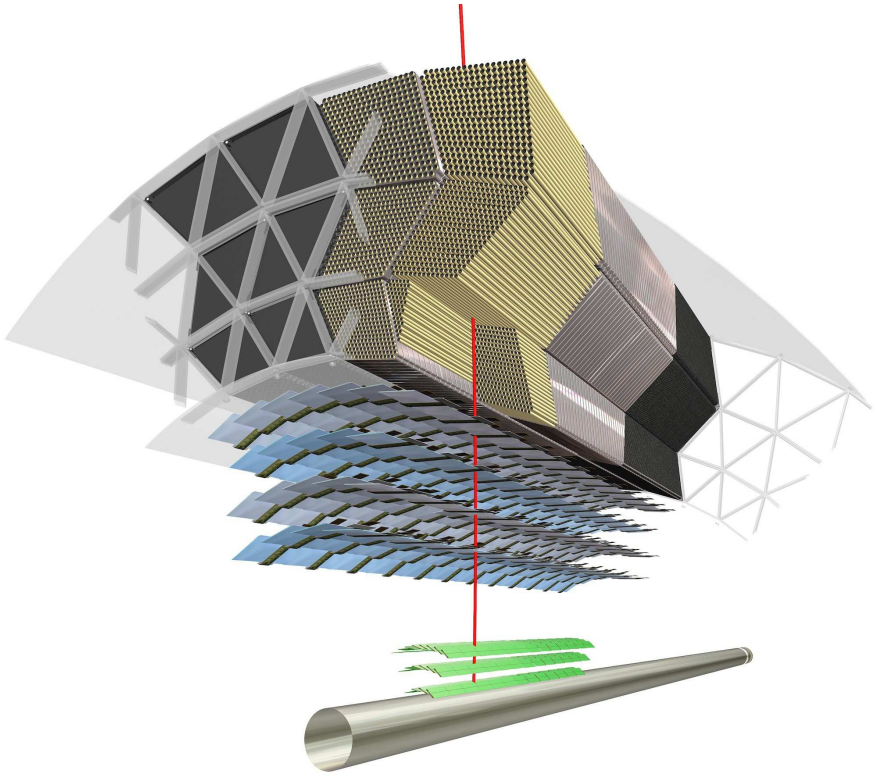


Figure 2.3 – Sketch of the ATLAS inner detector. It consists of three systems, from the innermost parts to the outermost: the pixel detector (in green), the silicon tracker (SCT in gray and blue) and the transition radiation tracker (TRT in yellow and black).

2.3.3 The ATLAS magnet system

The ATLAS magnet system consists of three sets of superconducting magnets [4].

- The solenoid (green in Figure 2.4) generates a uniform 2T field in z for the ATLAS inner detectors. It is 5.3m in length and 2.6m in diameter.
- The central barrel toroids (blue in Figure 2.4) consists of 8 each coil generating a 0.5T field in their central part for the muon system.
- The endcap toroids (red in Figure 2.4) are similar to the central barrel toroid providing a magnetic field for the muons system in the endcaps.

Due to the direction of the field produced by the toroids the charged particles are curved towards the front or the back of the detector, depending on their charge.

2.3.4 The ATLAS calorimeter

The ATLAS calorimeter system is shown in Figure 2.5. The calorimeters are located around the inner detector and the solenoid magnet. Their purpose is to measure the energy of the particles. They can detect photons, electrons and neutral hadrons, which do not leave tracks in the inner detector. It can be divided in three sections:

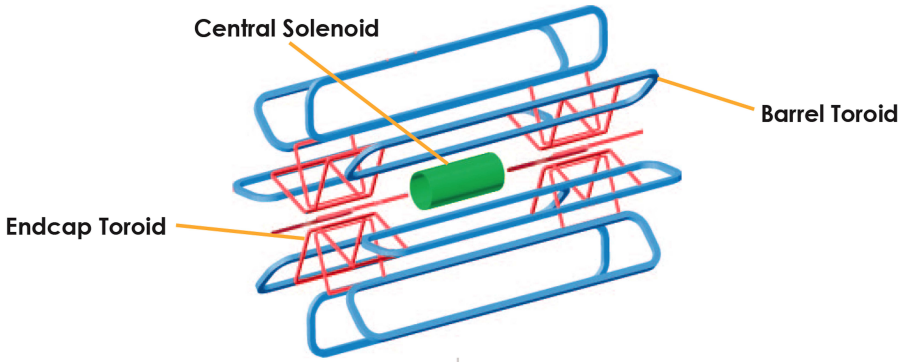


Figure 2.4 – Schematic view of the ATLAS magnet system.

the electromagnetic calorimeter (ECAL), the hadronic calorimeter (HCAL) and the forward calorimeter [5, 6] .

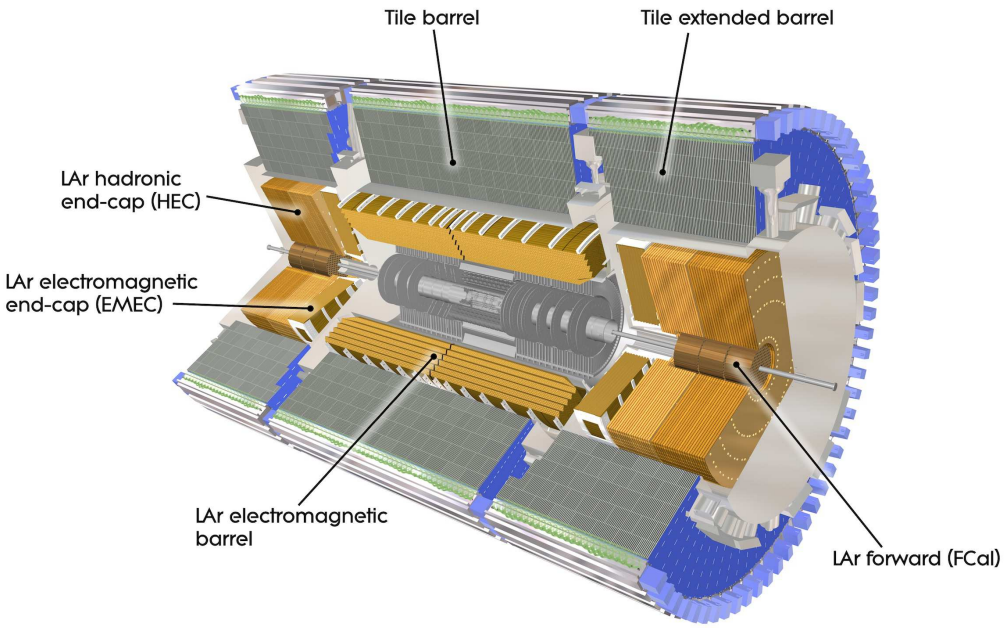


Figure 2.5 – Cut-away view of the ATLAS calorimeter system.

2.3.4.1 The electromagnetic calorimeter

The central electromagnetic calorimeter covers a region corresponding to $|\eta| \leq 1.475$, while the endcaps cover the region $1.375 \leq |\eta| \leq 3.2$. The absorbing material used is lead. Absorbing layers are alternated with active regions of liquid argon. The total thickness of the electromagnetic calorimeter is 22 radiation lengths in the barrel and 24 radiation lengths in the end-caps. This ensures that the energy leakage of the electromagnetic cascades outside of the ECAL is very small.

2.3.4.2 The hadronic calorimeter

The electromagnetic calorimeter is surrounded by the hadronic calorimeter. Hadrons are much heavier than electrons and their electromagnetic interaction with matter is much weaker. On the other hand, they interact strongly with atomic nuclei. This leads to hadronic cascades, which need much more material for the shower to develop. The hadronic calorimeter is divided in the hadronic barrel (TileCal) covering $|\eta| \leq 1.7$, the hadronic end-cap covering $1.5 \leq |\eta| \leq 3.2$ and a forward calorimeter (FCAL) covering $3.1 \leq |\eta| \leq 4.9$.

The tile calorimeter is a sampling calorimeter using iron as the absorber and plastic scintillating tiles as the active medium. The forward calorimeter (FCAL), in fact, acts both as a electromagnetic and hadronic calorimeter. Being in the forward region it is exposed to a high radiation flux due to the beam remnants emitted from the interaction point. The absorbing planes are made of copper and tungsten and the active part is liquid argon.

2.3.5 The ATLAS muon spectrometer

The outmost layer of the ATLAS detector is the muon spectrometer [7]. The ATLAS muon system is shown in Figure 2.6. Muons are relatively heavy particles and do not interact strongly, therefore they are usually not stopped in the calorimeters like other particles. An additional detector, the muon spectrometer, is used to identify them. It allows the determination of their momentum by measuring their tracks in the toroid magnetic field very precisely. The tracking is done using the Monitored Drift Tubes (MDT) in the central part and using Cathode Strip Chambers (CSC) in the forward part of ATLAS. The trigger is done separately by the Resistive Plate Chambers (RPC) and the Thin Gap Chambers (TGC). In total the ATLAS muon system covers the region where $|\eta| \leq 2.7$.

2.3.6 The ATLAS trigger system

The goal of the trigger system is to select and store the physics event of interest detected in ATLAS. At the LHC bunch crossing can occur every 25 ns, producing 1 petabyte of raw data per second in the detector. It would be impossible to store all events. Therefore a trigger system, made of three levels, filters successively all events, keeping only the ones potentially interesting [8]. The level 1 trigger, uses electronics located inside the detector, while the other two run primarily on a large computer cluster near the detector.

The *level 1* trigger works on a subset of information from the calorimeter and the muon system. The decision is taken in 2 μs, including the propagation delays in cables between the detector and the underground counting room, where the trigger logic is housed. All information from the detector must be stored in pipeline memories until the level 1 trigger decision is available. The maximum event rate after level 1 is $\sim$ 100 kHz.

The *level 2* trigger, refines the selection of candidate objects using all sub-detectors, including the inner tracker, which is not used at level 1. The data is accessed selec-

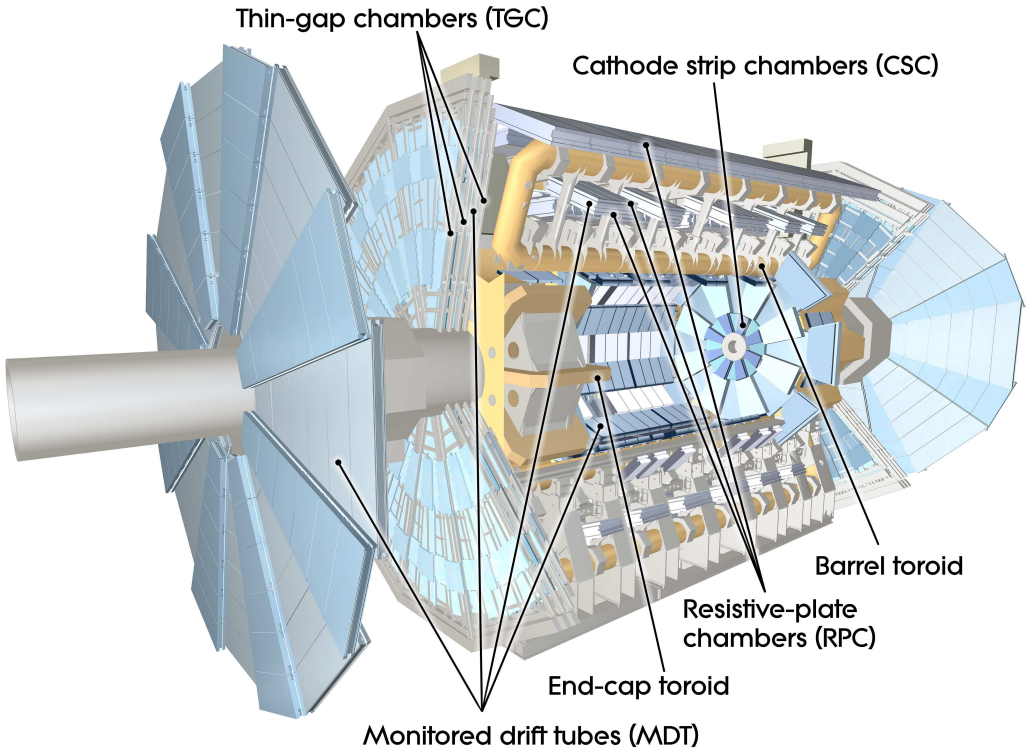


Figure 2.6 – Cut-away view of the ATLAS muon system.

tively in regions of interest defined by the *level 1* trigger. The trigger decision at *level 2* is taken in $40\mu s$ and reduces the event rate to $\sim$ 1 kHz.

The *level 3* trigger, uses off-line reconstruction algorithms that utilize all information from the event and taking into account calibration corrections. The decision time is 4 s per event.

2.4 Conclusion

The ATLAS detector is a general purpose detector, which allows myriad of physics measurements. The following list summarizes the overall detection capabilities of the ATLAS detector:

- Reconstruction of charged particles tracks is possible down to $|\eta| \leq 2.5$.
- Muon detection can be performed in a slightly larger range $|\eta| \leq 2.7$.
- The largest coverage is provided by the calorimeters $|\eta| \leq 4.9$.

Bibliography

[1] Christiane Lefevre. *LHC: the guide (English version)*. Technical report, 2009.

[2] ATLAS Collaboration, A Airapetian, Vladimir Cindro, Andrej Filipčič, Gregor Kramberger, Marko Mikuž, Matevž Tadel, and Dejan Žontar. *Inner Detector: Technical Design Report*. ATLAS Inner Detector Community, 1997.

[3] G Aad, M Ackers, FA Alberti, M Aleppo, G Alimonti, J Alonso, EC Anderssen, A Andreani, A Andreazza, JF Arguin, et al. *ATLAS pixel detector electronics and sensors*. *Journal of Instrumentation*, 3(07):P07007, 2008.

[4] HJ Herman. *The superconducting magnet system for the ATLAS detector at CERN*. *Applied Superconductivity, IEEE Transactions on*, 10(1):347–352, 2000.

[5] ATLAS collaboration et al. *ATLAS liquid argon calorimeter Technical Design Report*. *CERN/LHCC*, 96:41, 1996.

[6] ATLAS Collaboration et al. *ATLAS tile calorimeter: Technical design report*. *CERN/LHCC*, pages 96–42, 1996.

[7] ATLAS Muon Collaboration et al. *ATLAS muon spectrometer technical design report*. *CERN/LHCC*, 9:7–22, 1997.

[8] Georges Aad, B Abbott, J Abdallah, AA Abdelalim, A Abdesselam, O Abdinov, B Abi, M Abolins, H Abramowicz, H Abreu, et al. *Performance of the ATLAS Trigger System in 2010*. *The European Physical Journal C*, 72(1):1–61, 2012.

3
ATLAS Forward Proton

Contents

3.1 Introduction

The ATLAS Forward Proton (AFP) proposal intends to explore the nature of events where protons are intact in the final state[1] in the ATLAS detector [1]. This allows study of a new range of physics processes - typically elastic or diffractive scattering - where the proton radiates a virtual colorless object called Pomeron, or a photon, which will be studied here. The protons remain intact after the interaction and have a lower momentum, with respect to the beam. The LHC magnets along the beam can be used as a spectrometer to bend these protons and allow their detection in dedicated detectors situated close to the beam, far away from the ATLAS interaction point. The idea of the AFP project is to install such detectors at about 210 m from the interaction point. Tracking the protons through the LHC magnets and measuring their momentum permits analysis of the nature of the interaction. In addition, the AFP detector will be fully integrated with the main detector system, so that decay products of the object produced diffractively can be measured in the ATLAS detector. This physics has never been studied previously at such center of mass energies. Measurements performed with AFP would therefore be unique and potentially lead to new physic discoveries.

[1] Inelastic events only since AFP acceptance to elastic events is negligible

In Section 3.2 we will present the AFP detector and its objectives. In Section 3.3 we will present the detector installation within the LHC system. In Section 3.4 we will present the tracking detector for AFP and in Section 3.5 the timing detector.

3.2 AFP detector

The objective of the AFP detector is to measure the position, angle and time-of-flight from the interaction point of protons close to the LHC beam. The kinematic variables used to described diffracted protons are: the fractional momentum loss ξ

$$\xi = \frac{|\overrightarrow{p_b}| - |\overrightarrow{p}|}{|\overrightarrow{p_b}|} \tag{3.1}$$

the momentum transfer squared t:

$$t = (p_b - p)^2 \approx -p_T^2 \tag{3.2}$$

of the scattered protons, with p_b the momentum of the proton in the beam and p the momentum of the diffracted protons. The momentum transfer squared is approximately equal to the proton transverse momentum when ξ is small (proton scattered at small angles from the beam). If ξ_1 and ξ_2 are the two fractional momentum losses measured in the AFP detectors, we can determine the diffractive mass W via the relation:

$$W = \sqrt{s\xi_1\xi_2} \tag{3.3}$$

where $\sqrt{s}$ is the center-of-mass beam energy (8 TeV in present LHC runs - 14 or 13 TeV when the detector will be installed). This holds in the limit where $W \ll m_p$, *i.e.* the proton mass can be neglected. The range of protons that can be detected can be expressed in terms of these kinematic variables and for a forward detector at 210 meters from the ATLAS interaction point, the typical ξ of the measured protons spans:

$$0.015 \leq \xi \leq 0.15 \tag{3.4}$$

where the minimum obtainable value, here 0.015, is dictated by how close we can bring the detector to the proton beam (15 σ to the beam transverse size).

Approaching the detector closer results in a smaller ξ acceptance, but could result in perturbation or even destruction of the LHC beam. The maximum reachable value, here 0.15, is usually set by the acceptance of the upstream beam collimators, designed to absorb scattered particles, so as to prevent the superconducting magnets from quenching.

The constraints due to the LHC machine on the range of the ξ of the diffracted protons are illustrated in Figure 3.1. We can see the spectrometer effect of the LHC dipole magnets: going further away from the ATLAS interaction point allows detection of protons with lower ξ.

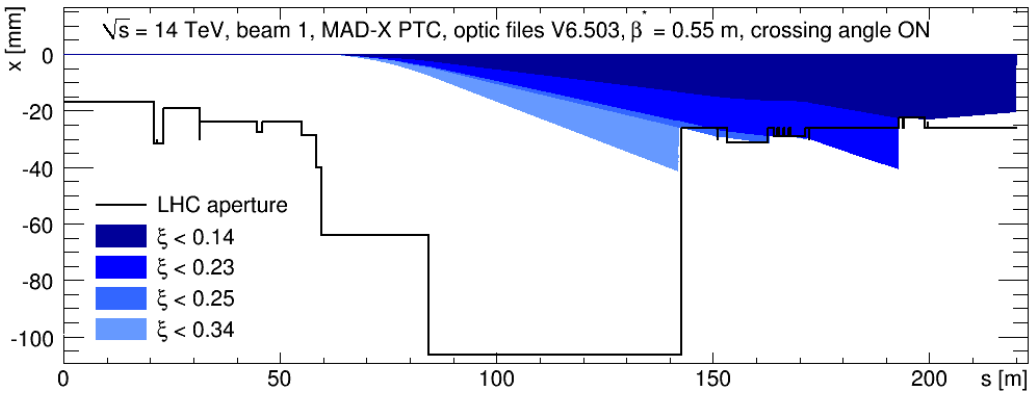


Figure 3.1 – Simulation of the proton transport from the interaction point to the location of the AFP detectors, showing the limitation at high ξ due to LHC transport.

3.3 Detector deployment

The configuration of the LHC beamline in the ATLAS forward region is shown schematically in Figure 3.2. This project proposes to install detectors in the region located approximately 210 meters from the IP1 interaction point. The detectors would be installed in both beamlines situated downstream from the central detector. At these points, due to their momentum loss in the collision, each intact proton will emerge from the beam envelope to be detected. Since the protons emerging from the beam at this distance have a very small elevation angle two separated tracking stations are required (in order to precisely measure their position). The proposal is to install them at 206 meters and 214 meters from IP1, as seen in Figure 3.2.

In the detectors, the acceptance (or range of detectable particles) and ultimate experimental resolution, is dictated by the LHC beam optics. In reality, the acceptance is limited also by the total active area of the detector, its spatial resolution, and the minimum acceptable distance of the detector to the beam.

3.3.1 Detector acceptance

After transport through the LHC magnets, from the point of interaction to the location of the AFP detector at 206m, the simulated scattered proton density in the transverse plane is shown in Figure 3.3 [2]. As it can be seen, the density is highly non-uniform with a higher density very close to the beam (lower ξ protons). This means that the AFP detectors will be required to come very close to the beam in order measure them - typically from 2 to 3 mm to the beam center. This is a challenge as it must be done without perturbing the LHC beam.

Figure 3.3 illustrates two important points:

- The hit positions are tightly packed requiring micrometric resolution on the proton position for a precise ξ measurement of the protons. The requirements on the spatial resolution of the AFP tracker are set to 10μm along the x-axis and 30μm in the y-axis direction. An active area of 20 by 20 mm, is sufficient to capture most of the protons that have interacted.

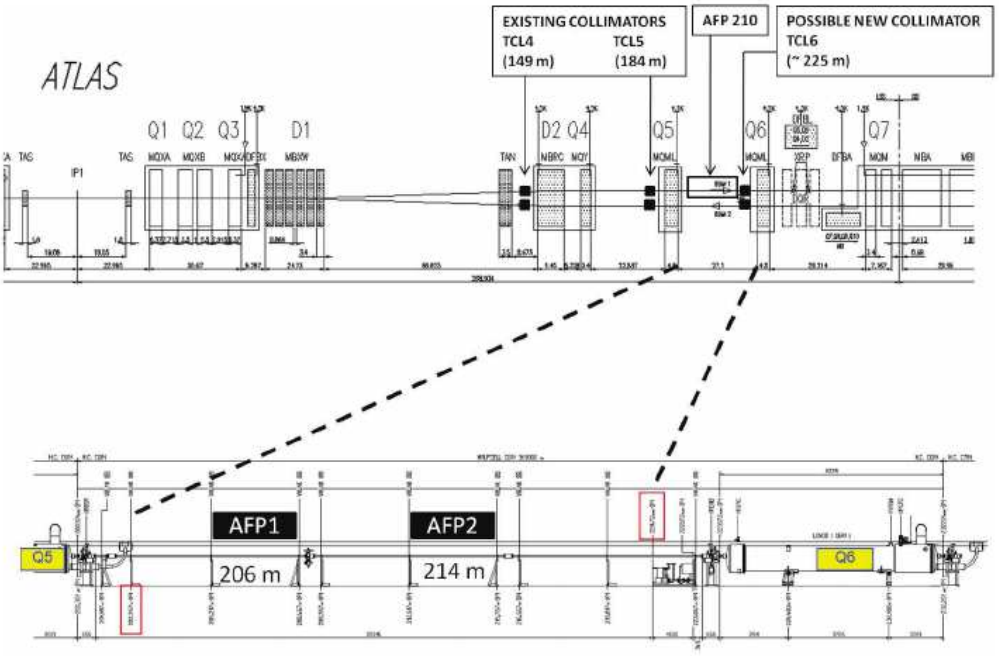


Figure 3.2 – Schematic view of the LHC beamline from the interaction point IP1 to the 210 meter region where the AFP detectors installation is proposed. The proposal is to distribute the AFP detectors on two location situated at 206 meters and 214 meters from the interaction point.

- The hit density is generally greater than one in most of the acceptance region (multiple hits regions). This effect is due to the large multiplicity of primary vertices in the central detector ($\mu \geq 1$, see Figure 1.3), and requires an additional mean of proton identification in the case of multiple hits in the AFP detector. The solution chosen is to add a time-of-flight detector to the AFP tracker, which matches each diffracted protons to their vertices of origin, by measuring their travel time. Two protons originating from different vertexes will arrive at slightly different times in AFP. The precision required here is on the order of 10 picoseconds for a 3 mm resolution on the vertex position.

3.3.2 Detector structure

According to the AFP detector requirements, a set of two detectors are proposed: a position-tracking detector, similar to the ATLAS inner tracking detectors (see Section 2.3) and a time-of-flight detector. The detectors are located in two stations near the LHC beampipe, as shown in Figure 3.4.

In order to be able to put tracking detectors as close as 2.5 millimeters from the LHC beam, the beampipe at the location of the detector must be modified in order to accommodate this tracking and timing measurement. Two solutions are proposed: a movable beampipe and a "roman pot" structure.

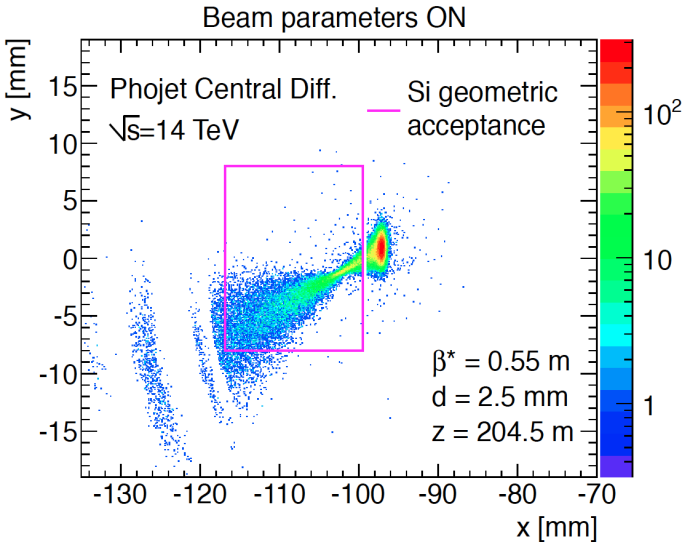


Figure 3.3 – Plot of the diffracted proton hit map in the xy-plane (perpendicular to the beam) showing the proton density at 206 meters. The size and position of the AFP detector is indicated as the pink box on the plot, covering the region of 2 by 1.9cm at 2.5mm from the LHC beam (region of highest density in red).

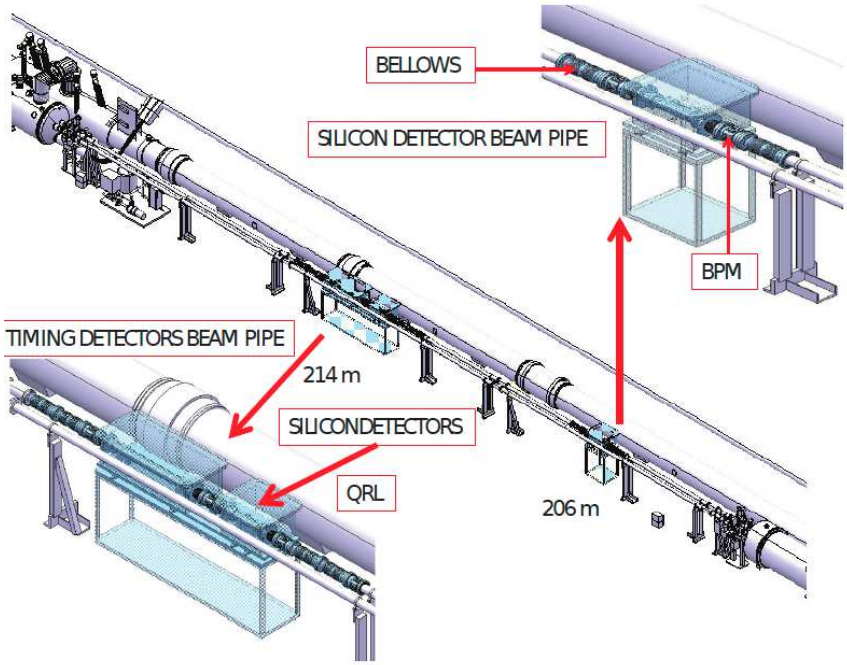


Figure 3.4 – Diagram showing the location of the insertion of the AFP detectors in the LHC beampipes at 206 and 214 meters.

3.3.3 The movable beam pipe

The principle of the movable beam pipe is illustrated in Figure 3.5. In order to get the detector close to the beam, part of the LHC beam pipe is replaced by a movable beampipe. The geometry of the pipe inserted is modified from the regular cylindrical shape to have a vertical, flat (floor) section where the AFP detectors are positioned, and which can be brought close to the LHC beam, moving the entire pipe horizontally. To do so, it is attached to the main LHC pipe via bellows, and sits on an x-y table.

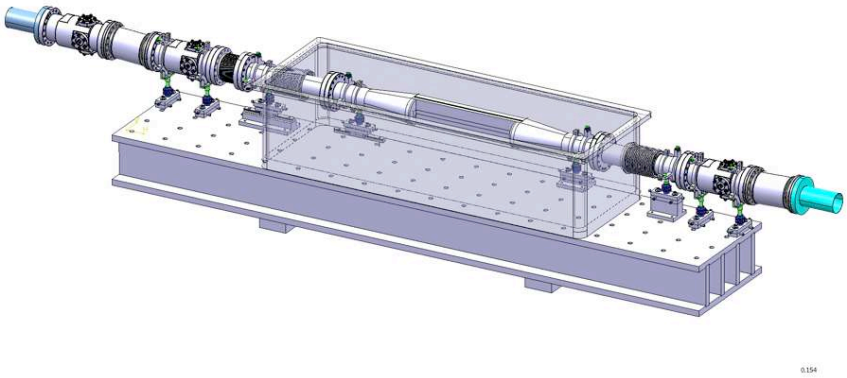


Figure 3.5 – Diagram showing the principle of the movable beampipe insertion in the LHC. Part of the LHC pipe is replaced by a movable section. The geometry of the pipe inserted is modified from the cylindrical shape to have a vertical, flat section called a floor, which can be brought close to the LHC beam.

The AFP detectors are located inside a box on the x-y table, with their sensitive region positioned as close as possible to the "floor" of the movable beam pipe, in order to capture as many protons as possible in the low ξ region. A detailed illustration of the movable beampipe is shown in Figure 3.6, where the flat floor of the pipe next to which the detectors will be placed is clearly visible. The beampipe is connected to the rest of the LHC apparatus with bellows that are not shown here.

The advantage of the movable beampipe technology is to allow for large movable beampipe lenght (≈ 1 meter) and therefore can accommodate large detectors systems. However due to the change of the section geometry that they impose on the beampipe, they require special attention to match the impedance of the rest of the LHC ring.

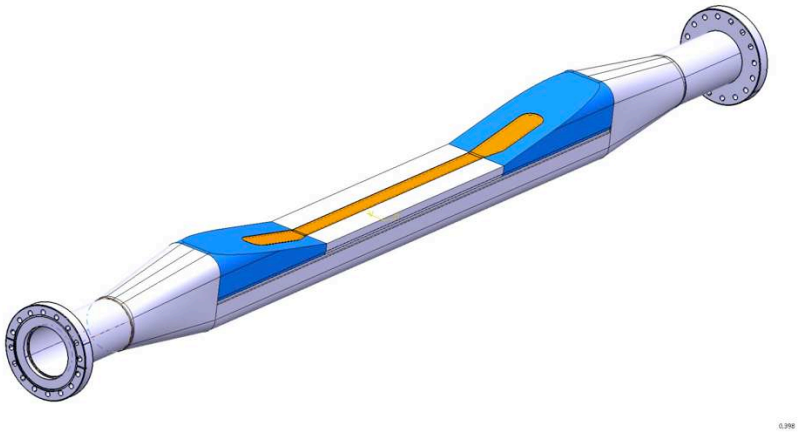


Figure 3.6 – Sketch showing a detailed view of the movable beampipe, for use in the AFP experiment. The LHC beam circulates inside the inner part of the tube, while protons scattered at small angles emerge from the tube crossing the "window". Most of them come out in the orange section which is made of thinner material in order to reduce parasitic interactions with the beampipe and are detected in the AFP detectors (not shown here) sitting on the tube flat "floor". A particular taper angle (blue part) is designed to limit the impedance mismatch and the RF pickup of the movable beampipe, at the cost of more dead material.

3.3.4 The "Roman pot"

The principle behind the Roman pot design is very similar to the movable beampipe. However, the idea here is to move the detector only, as opposed to the movable beampipe option, where the integrated detector and beampipes are moving. Due to its small spatial extension the "Roman pot" design does not suffer as much of impedance mismatch. The "Roman pot" was designed for the Totem and ALFA experiments and has been installed in the LHC, it benefits from a strong user experience [3].

A diagram of the Roman pot design is shown in Figure 3.7. In this design the tubular geometry of the LHC beampipe is preserved (shown in gold color), but a movable pocket (in blue) with a detector inside can be brought close to the beam to measure the diffracted proton.

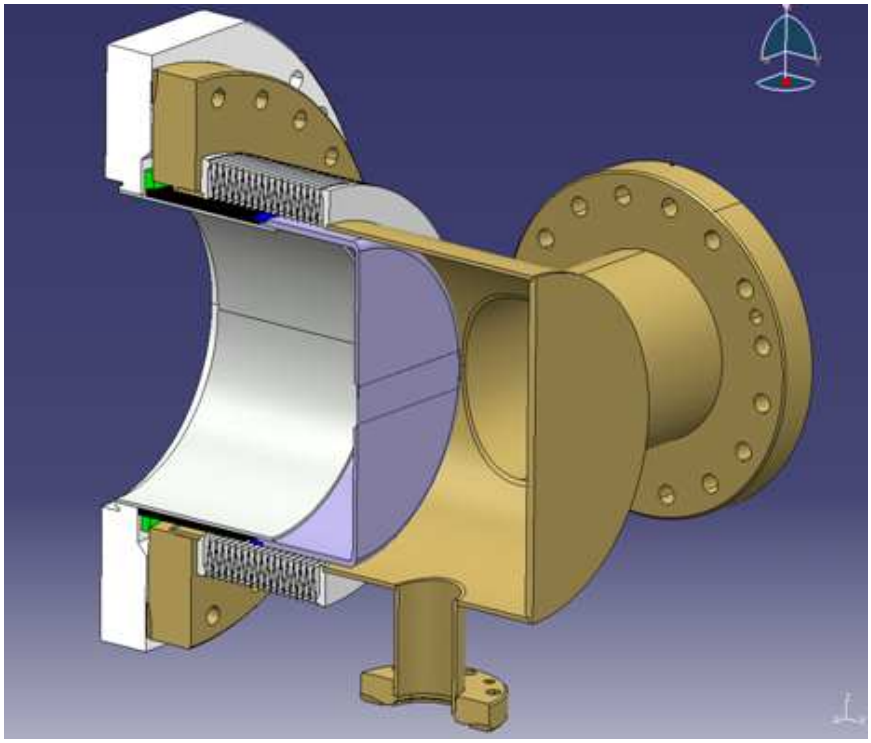


Figure 3.7 – Diagram of the "Roman pot" design concept. Here the position of pipe is fixed (in gold color) but a movable pocket (in blue), with a detector inside, can be brought close to the beam to measure the diffracted protons.

The advantage of the "Roman pot" option is the simplicity and robustness of the design. However, the small size of the detector pocket prohibits the installation of large or bulky detectors. It should be noted that the Totem and ALFA experiments have already installed Roman pots on the LHC beam pipe, and they have been proven to be completely functional.

3.4 The tracking detector

The purpose of the tracking detector is to provide a measurement of points along the trajectory and hereby the ξ value of the diffracted protons. The area to be covered by the tracker is approximatively 20 by 20 mm. To do so, it was decided to use Si pixel detectors instrumented with the FE-I4 readout integrated circuit, originally designed for the Insertable B-Layer (IBL) upgrade of the ATLAS pixel detector [4]. This circuit is made of 26880 $50 \times 250\mu$m size pixels. There are multiple advantages of using the FE-I4 circuit for the AFP tracker:

- To measure protons at low ξ, the AFP tracking detector must be able to get close to the beam. The closest approach possible, avoiding beam perturbations, is set at about 15 σ corresponding to 3.5 mm from the beam center. An advantage of the FE-I4 circuit is that it has a very small dead zone: ≤ 100 μm (corresponding to the inactive edge of the chip), allowing the immediate reach of very low ξ protons (minimal dead-zone).
- The FE-I4 chip has also been designed to survive the high radiation environment that exists in the acceptance region of the AFP detector.
- The FE-I4 provides an excellent x-y position tracking ($50\mu m \times 250\mu m$ pixels) and meets the 10 μm position and 2 μ rad in angle precision requirements. Two assemblies of 5 layers each of FE-I4 chips are used.

A diagram showing the assembly of 5 FE-I4 chips with Silicium detectors for the AFP is given in Figure 3.8. A metal frame provides support and a fixed position - stable at a few microns level - for the 5 layers. Due to the high power consumption of these ASICs a cooling system is also required to remove the heat generated. This tracking system is compact and therefore could be integrated either in the "Roman pot" or the movable beampipe options.

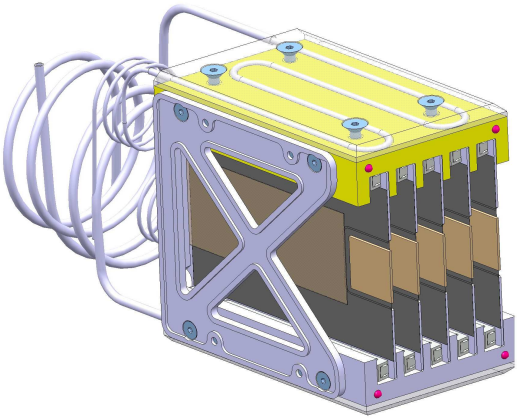


Figure 3.8 – Diagram of the AFP tracking detector. It consists of the assembly of 5 individual silicon pixel detectors.

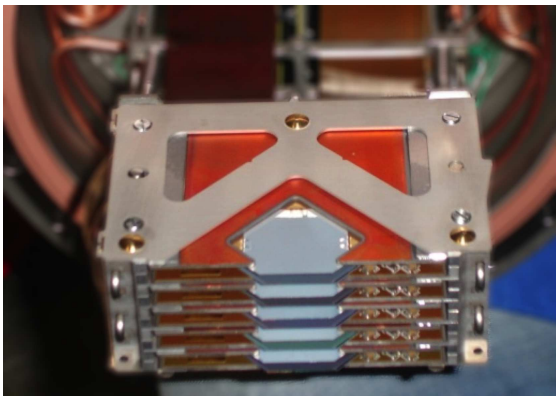


Figure 3.9 – Photograph of the tracking detector used in the Totem experiment.

3.5 The timing detector

The timing detector in AFP is required to suppress pile-up events (reject protons coming from overlap events). In order to preserve timing information at a few picoseconds level, signal detection from the proton uses prompt Cerenkov radiation created in a succession of quartz bar [5]. This process is illustrated in Figure 3.10

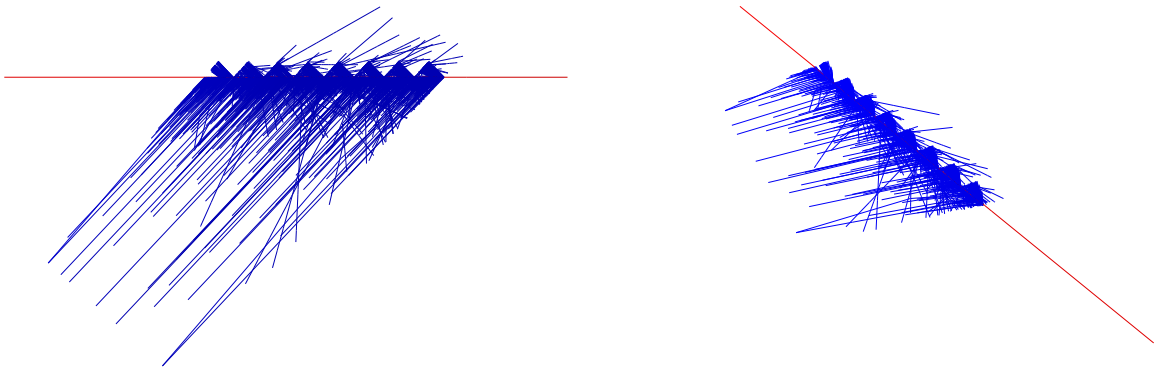


Figure 3.10 – Illustration of the Cerenkov photon generation (in blue) in quartz bars by a diffracted proton (in red). Some photons are created at the transition region and emitted outside the bars.[6]

In order to measure precisely the time of arrival of the Cerenkov photons created, a very fast photo-detector must be used. Also, some spatial segmentation is required in order to be able to cope with multiple protons emission at various ξ. To do so, a micro-channel plate photo-detector is the perfect candidate (as will be shown in Chapter 5). This principle is shown in Figure 3.11, where a diagram shows the integration of the quartz bars (in light green) with the micro-channel plate photo-detector (in gray), collecting light from the bars. Several rows of bars are used to provide a few mm position resolution on the protons. Due to non-uniform proton distribution in the bars, they are made increasingly smaller closer to the beam, to cope with higher rates.

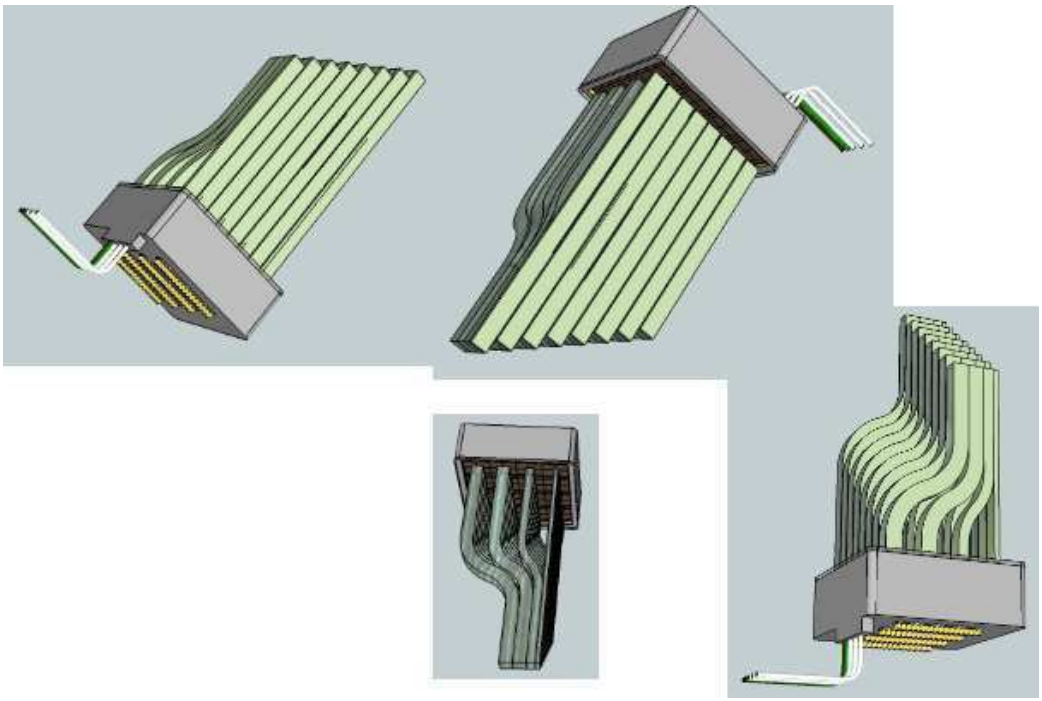


Figure 3.11 – Diagram illustrating the integration of the quartz bars (in light green) with the micro-channel plate photo-detector (in gray) collecting the light from the bars. Several rows of bars are used in order to have a few 10μm position resolution of the protons. Due to non-uniform proton distribution, the bars are made increasingly smaller closer to the beam, in order to cope with the higher rates.

Bibliography

[1] ATLAS Collaboration. *Letter of Intent for the Phase-I Upgrade of the ATLAS Experiment*. Technical report, CERN-LHCC-2011-012, 2011.

[2] R Staszewski and J Chwastowski. *Transport simulation and diffractive event reconstruction at the LHC*. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 609(2):136–141, 2009.

[3] Giovanni Anelli, G Antchev, P Aspell, V Avati, MG Bagliesi, V Berardi, M Berretti, V Boccone, U Bottigli, M Bozzo, et al. *The Totem experiment at the CERN Large Hadron Collider*. *Journal of Instrumentation*, 3(08):S08007, 2008.

[4] Marlon Barbero. *The FE-I4 pixel readout chip and the IBL module*. Technical report, ATL-COM-UPGRADE-2011-024, 2012.

[5] MG Albrow, Heejong Kim, S Los, M Mazzillo, E Ramberg, A Ronzhin, V Samoylenko, H Wenzel, and A Zatserklyaniy. *Quartz Cherenkov counters for fast timing: QUARTIC*. *Journal of Instrumentation*, 7(10):P10027, 2012.

[6] Mateusz Dyndal. *Cerenkov photon generation in quartz bars by a diffracted proton.* Simulation work.

4

Search for anomalous coupling between photon and W boson at the LHC in ATLAS

Contents

4.1 Introduction

In this Chapter we present the prospect for sensitivity to quartic anomalous couplings between photons and W boson, which was done for the ATLAS experiment in the context of the installation of the AFP detectors in the forward region. The aim of this study is to investigate the potential of new physics discoveries when protons are left intact after interaction and are then detected in the AFP detector. This ATLAS Forward Physics project allows the extension of the physics reach of ATLAS, by enabling observation and measurement of a range of processes where one or both protons remains intact and which would be otherwise difficult or impossible to study [1]. This Chapter focuses on the search for anomalous quartic gauge couplings between the photon and the W boson ($\gamma\gamma WW$) [2].

4.2 Theoretical context

4.2.1 Diffractive physics

In typical LHC interactions the proton collisions are destructive, that is the protons do not exist in the final state. Indeed, the partons (quark, gluons), constituents of the protons and taking part in the interaction, are colored objects. Their interactions are ruled by Quantum Chromodynamics, the theory of strong interactions. A requirement of the strong interaction, called *confinement*, is that all physically observable particles carry no color charge. Therefore, color exchanges happening during the proton interaction by the mean of parton transfer, usually suppresse the color neutrality of the protons in the final state, and they are destroyed.

The decay products of the energetic protons are detected as hadronic activity, in the forward region of the detector. Diffractive physics, on the contrary, describes processes where the protons are scattered from the beam at small angles and remain intact in their final state [3].

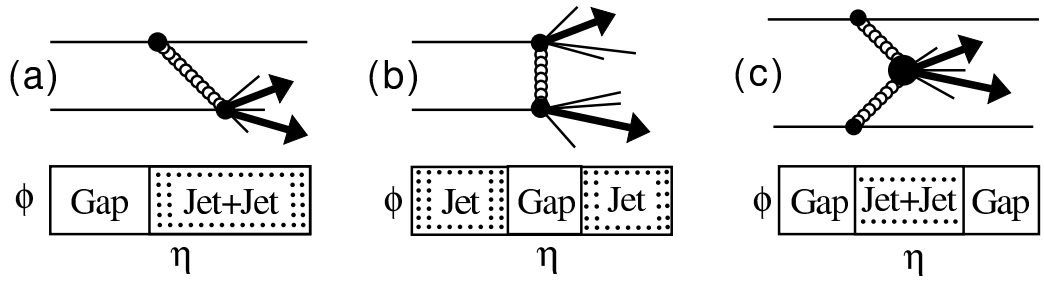


Figure 4.1 – Types of diffractive events in a hadronic collider: (a) single diffractive (SD), (b) double diffractive (DD) and (c) central exclusive (CE) [4]. The lower part of each Figure illustrates the geometric repartition of the activity in the detector versus the pseudo-rapidity η for each case. between no activity *gaps* and *jets*.

There has been evidence, first in UA8 and then in the HERA experiments, of events where strong interactions between two colliding particles leave no detectable trace into the forward region. The absence of energy, except for the noise, in that

region is described as a *rapidity gap*. In these events, the protons are simply scattered from the beam at small angles, and a hard event is produced in the central detector. This is explained theoretically by the exchange of colorless objects, called pomerons, during the interaction, leaving the protons intact in the final state.

There are three kinds of diffractive events (see Figure 4.1): single diffractive (SD), double diffractive (DD) or central exclusive (CE) events. In the first case, only one proton remains intact in the final state, the other one is destroyed. The momentum transferred during the interaction is small and the intact proton is scattered at small angle. In the second case, both protons are destroyed and a colorless object is exchanged between central objects, for instance appears as jets. And in the last case, central exclusive production, both protons survive the collision. These three class of event are characterized by three different types of measurement versus the pseudo-rapidity in the detectors. This is illustrated in Figure 4.1.

4.2.2 Photon induced processes

At the LHC, the protons can interact not only strongly but also by photon exchange. Since the photon is not a colored object, the protons can remain intact after the interaction. The process $pp \to pXp$, where the object X is produced by the two photon exchange; $\gamma\gamma \to X$, is illustrated in Figure 4.2.

4.2.3 Cross-section of interaction

The spectrum of photon production can be calculated in the Equivalent Photon Approximation or EPA [5] and gives for a photon of energy E_γ and virtuality Q^2 [6]:

$$dN = \frac{\alpha}{\pi}\frac{dE_\gamma}{E_\gamma}\frac{dQ^2}{Q^2}\left[\left(1-\frac{E_\gamma}{E}\right)\left(1-\frac{Q^2_{min}}{Q^2}\right)F_E + \frac{E^2_\gamma}{2E^2}F_M\right] \tag{4.1}$$

where E is the energy of the proton of mass m_p. $Q^2_{min} = m_p^2 E_\gamma^2/[E(E-E_\gamma)]$ is the minimum virtuality possible of the photon given the kinematic parameters and F_E and F_M are the electric and magnetic form factors. To obtain the production cross-section, the photon fluxes are integrated over Q^2:

$$f(E_\gamma) = \int_{Q^2_{min}}^{Q^2_{max}} \frac{dN}{dE_{gamma}dQ^2}dQ^2 \tag{4.2}$$

By integrating the product of the two photon fluxes $f(E_{\gamma 2}) \times f(E_{\gamma 1})$ while keeping the two-photon invariant mass $W = \sqrt{E_{\gamma 1}E_{\gamma 2}}$ constant, the two-photon effective luminosity spectrum $dL^{\gamma\gamma}/dW$ is obtained. The shape of $\frac{dL^{\gamma\gamma}}{dW}$ is shown in Figure 4.3. The production of heavy objects extending in a clean diffractive environment is of particular interest at the LHC. The production rate of massive objects is, however, limited by the decrease of photon luminosity at high invariant mass W. The integrated two-photon luminosity above $W \geq W_0$ for $W_0 = 23$GeV, $2 \times m_W \approx 160$GeV, and 1TeV is respectively 1%, 0.15% and 0.007% of the luminosity integrated over the whole mass spectrum.

And the differential cross-section $\frac{d\sigma}{d\Omega}$ for a photon induced $pp \to pXp$ process is given by:

$$\frac{d\sigma}{d\Omega} = \int \frac{d\sigma_{\gamma\gamma\to X}(W)}{d\Omega} \frac{dL^{\gamma\gamma}}{dW} dW \tag{4.3}$$

where $d\sigma_{\gamma\gamma\to X}/d\Omega$ is the differential cross section of the process $\gamma\gamma \to X$, versus the invariant mass W of the two-photon system.

4.2.4 W pair production by photon exchange

For this work we have studied the photon-production processes $\gamma\gamma \to W^+W^-$, their Feynman diagrams are shown in Figure 4.4.

In the standard model the Lagrangian describing these interactions can be written as:

$$\mathcal{L}_{WW\gamma} = -ie(A_\mu W^-_\nu \overleftrightarrow{\partial^\mu} W^{+\nu} + W^-_\mu W^+_\nu \overleftrightarrow{\partial^\mu} A^\nu + W^+_\mu A_\nu \overleftrightarrow{\partial^\mu} W^{-\nu}) \tag{4.4}$$

$$\mathcal{L}_{WW\gamma\gamma} = -e^2(W^-_\mu W^{+\nu} A_\nu A^\nu - W^-_\mu A^\mu W^+_\nu A^\nu) \tag{4.5}$$

where the asymmetric derivative is $X \overleftrightarrow{\partial^\mu} Y = X\partial^\mu Y - Y\partial^\mu X$. We can identify in the Lagrangians the trilinear and quartic coupling terms (Figure 4.4) present in the Standard Model.

4.3 $pp \to pWWp$ observation

The $\gamma\gamma$ total luminosity is shown in Figure 4.3, which leads to a cross section of the exclusive process $pp \to pWWp$ of 95.6fb, for the Standard Model. This value has been corrected for the survival probability factor, which is 0.9 at the LHC.

This cross section is rather small, *i.e.* three order of magnitudes lower than the inelastic production produced by quark-anti-quark annihilation $q\overline{q} \to W^+W^-$ ($\sim 95\%$) and also via gluon-gluon fusion $gg \to W^+W^-$ ($\sim 5\%$) totaling 113.6 pb. Therefore, a substantial amount of luminosity has to be collected to get a significant WW sample. This can only be achieved at the LHC running at high instantaneous luminosity $\mathcal{L} = 10^{32} - 10^{34} cm^{-2}s^{-1}$. Under such conditions the two-photon events must be selected using a forward proton-tagging detector such as the one described in Chapter 3. In fact, for the rest of this study the acceptance in the momentum fraction loss ξ is assumed to be $0.015 \leq \xi \leq 0.15$, corresponding to the acceptance of the proton tracking detectors installed at 210 meters from the ATLAS interaction point.

As presented in Chapter 2, at present luminosity, the number of interactions occurring at the same time is in the order of 32 - however, at higher luminosities it can increase to 50 or even 100 after 2015. The consequence is that two protons from 2 single diffractive minimum bias events can give a hit in both forward detectors, while a third standard inelastic event could mimic the two WW signal in the central

detector. In this case, protons detected in the forward detectors are not related at all to the hard event producing the W pair. To reject this type of background, timing detectors are used. They measure the time of arrival of the two protons with a 5-10 ps precision, constraining the vertex position from which the protons come within $10ps/\sqrt{2} \times c = 2.1mm$ (where c is the velocity of the scattered protons, and close to the speed of light). Matching this information to the reconstructed vertex position determined by the inner tracker (see Chapter 2), the overlaid background can be suppressed by a factor 40 [7]. This is illustrated in Figure 4.5.

4.3.1 *W* boson decay

A W boson can decay hadronically into a pair of quarks, $q\overline{q}$ ($\sim 68\%$) or leptonically into a lepton and its neutrino, $l\nu_l$ ($\sim 32\%$). The final state is formed by four fermions, which are the decay products of the two W bosons. There are three possible configurations, fully hadronic final state, when both W's decay hadronically, semi-leptonic, when one W decays into $l\nu_l$ and the other in $q\overline{q}$ and the leptonic case. Then finally, when both W's decay into a lepton and its neutrino. The branching fractions of the three final states are:

- Hadronic: 45.6%
- Semi-Leptonic: 43.9%
- Leptonic: 10.5% [1]

In this thesis the semi-leptonic decay of the W pair was considered: one of the W decaying into a lepton and neutrino and the other into a $q\overline{q}$ pair, which is seen experimentally as two hadronic jets. This was performed as an addition to the leptonic-only decay study in order to include more statistics for a possible improvement of the sensitivity over anomalous couplings.

4.3.1.1 Leptonic case

Preliminary studies showed [6, 8] that using AFP detectors, the sensitivities on anomalous coupling parameters can be improved by 4 orders of magnitude with respect to LEP results using $WW \to ll\nu\nu$, with $l = e, \mu$. However, these studies were performed without considering the multiple proton-proton background - which is also the case for the present study. Therefore, a complete analysis with fully simulated events was performed, under two harsh pile-up conditions: $\mu = 23$ and $\mu = 46$, and for two expected integrated luminosities of $\mathcal{L} = 40fb^{-1}$ and $\mathcal{L} = 300fb^{-1}$. Apart from simulating the exclusive QED anomalous signal, a full set of background processes was considered ('direct' from double tagged events such as two-photon dileptons, WW production in DPE, etc. as well as 'pile-up', for instance, Drell-Yan, WW, $t\bar{t}$, etc.). With the timing information alone, it is difficult to reduce the combinatorial background. Dedicated, but rather simple cuts are needed in addition. Even at high

[1] But we only consider final state in electrons and muons

pile-up, the number of reconstructed tracks with $p_T \geq 400MeV$ associated to the primary vertex, shows large discriminating power. Moreover, the requirement of a high mass to be reconstructed in the AFP detector, that is, $\sqrt{s\xi_1\xi_2}$, greatly reduces the expected background. The study showed that even with high pile-up such analysis is feasible and leads to similar sensitivites as the ones obtained using the fast simulation.

4.3.2 Semi-leptonic signal and background

The semi-leptonic decay of the W pair can be written as:

$$\gamma\gamma \to W^+W^- \to l\nu q\overline{q} \tag{4.6}$$

Several background processes could mimic this signature, which has two intact protons in the final state. They are the following:

- $DPE \to WW \to l\nu q\overline{q}$ - W pair production by double pomeron exchange.
- $DPE \to W \to q\overline{q}$ - single boson production, the dijet system is misreconstructed as a lepton plus dijet system.
- $W + X \to W + jet$ - single boson plus jet production. The final state is misreconstructed as a lepton plus dijet system.
- $DPE \to t\bar{t}$ - $t\bar{t}$ pair production where the decay products are misreconstructed as a lepton plus two jets.

  Additionally full leptonic backgrounds were also included in the production, in order to study the semi-leptonic plus leptonic case. Both backgrounds should therefore have small contributions in the semi-leptonic case.
- $DPE \to l\bar{l}$ - two-photon dilepton production by double pomeron exchange, where one of the lepton is misreconstructed as a jet.
- $\gamma\gamma \to l\bar{l}$ - dilepton pair production by photon exchange, one of the lepton is misreconstructed as a jet.

In this analysis no pile-up contribution has been considered: we are only looking at diffractive background with two protons in the final state. In reality lots of pile up events, where the protons come from single-diffractive or elastic scattering and a hard object from another (pile-up) event, mimics the WW production by photon exchange. This analysis is a first study of the sensitivity of this signal in ATLAS using forward detectors.

4.4 Monte-Carlo samples

Due to the diffractive nature of the signal, the backgrounds to be considered are all other diffractive processes, where two intact protons are observed in the final state. The signal and backgrounds have been generated using the Forward Physics Monte Carlo (FPMC) [9]. FPMC is a monte carlo software designed to produce particles with one or two leading intact protons and some hard scale in the event.

4.4.1 Preselection

A preselection at the generator level is applied on all samples requiring a minimum p_T in hadronic jet production $PT_{MIN} = 20GeV$. Similarly, the minimum and maximum beam momentum loss, ξ_{min} and ξ_{max} are set at the generator level close to the final AFP acceptance: $YWWMIN = 0.001$ and $YWWMAX = 0.2$. This preselection is set in the FPMC datacards (parametrization files used for the Monte Carlo generation).

The event production and cross sections are reported in Table 4.1. Due to the large span of cross sections (from 20 pb to 0.3 fb), a large number of events had to be generated in order to get significant numbers at the end of the analysis. A total of 50 million events have been generated for each sample of higher cross section.

Table 4.1 – Total cross section for all considered process.

Process	Total cross section	Production flag
$DPE \to l\bar{l}$	6.7 pb	PTMIN = 20
$\gamma\gamma \to l\bar{l}$	0.29 fb	PTMIN = 20
$DPE \to WW$	5.98 fb	PTMIN = 20
$DPE \to W + X$	20.6 pb	PTMIN = 20
$W + jet$	4.28 pb	PTMIN = 20
$DPE \to t\bar{t}$	38.8 fb	PTMIN = 20

Subsequently, all signal and background events have been reconstructed using the Fast ATLAS simulator (ATLFast). ATLFast takes the physical properties of the generated particles (charge, momentum, η, ...) and simulates their detection and measurement in the ATLAS detector. For example, due to the limited η acceptance of ATLAS, some generated particles at high η would not be seen in the ATLAS detector, and therefore they are not reconstructed. To summarize, ATLFAst effectively applies all the ATLAS limitations in detection to the generated particles of each simulated event.

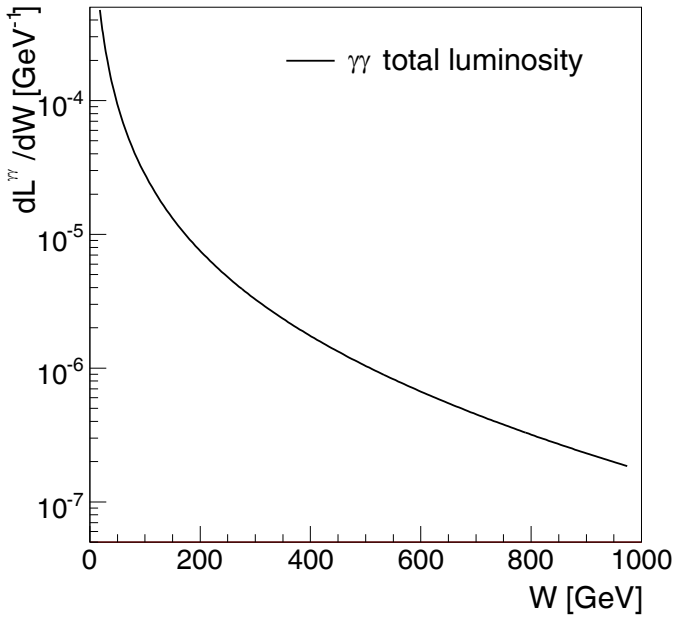


Figure 4.3 – Relative effective luminosity $\gamma\gamma$ in pp collision versus the invariant mass of the two photons. The maximum virtuality of the two photons is set at $Q^2_{max} = 2GeV^2$ [6].

Figure 4.4 – Feynman diagrams contributing to the process $\gamma\gamma \rightarrow W^+W^-$. The two diagrams exist in the standard model, the left one is built upon the trilinear coupling $WW\gamma$ while the right one uses the quartic coupling $WW\gamma\gamma$. Anomalous quartic couplings follow the diagram on the right.

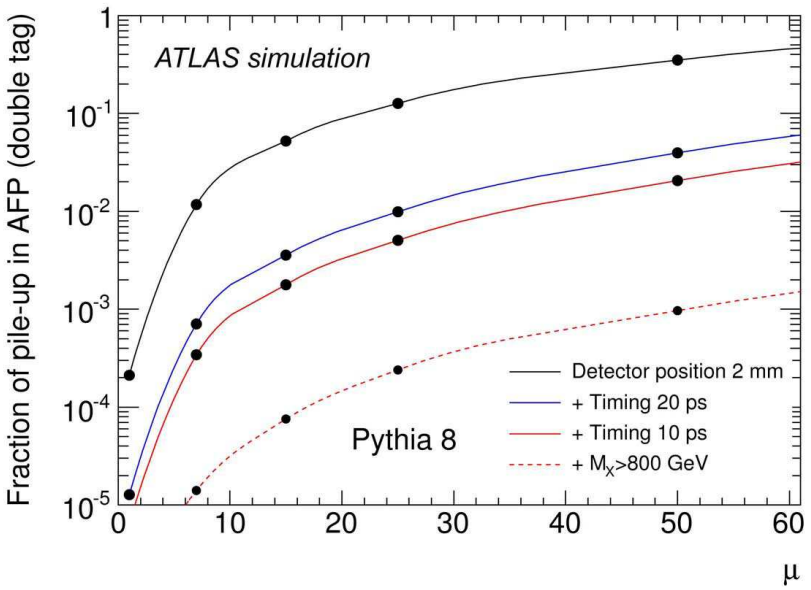


Figure 4.5 – Fraction of non-diffractive events that pass the successive experimental selection as a function of the average pile-up multiplicity μ. The experimental cuts are: double tag in the AFP detector (in black), timing cut at 20 picoseconds (in blue), 10 picoseconds (in red), and finally the missing mass M_X being greater than 800 GeV (dotted red line).

4.5 Signal selection

After production and reconstruction, signal and backgrounds events are stored into ROOT files and analyzed using a custom C++ code. The aim of the analysis is to select signal events from the background. Therefore, the code runs over all samples with a selection process based upon several *cuts*.

4.5.1 Signal and background representation

For the rest of the analysis the signal and background will be represented using the color code on the left of Figure 4.6.

On the right of the Figure 4.6 is the color code used in case of a display of the effect of successive cuts on one sample of the signal of the background.

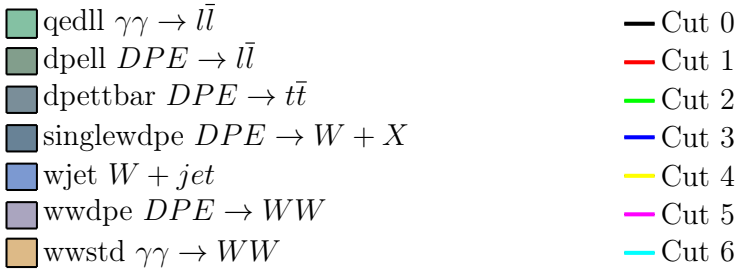


Figure 4.6 – Legend for the color of cuts applied to signal and background in this work with assigned colors (left) and cut numbers (right).

4.5.2 AFP acceptance cut

The first cut is the AFP acceptance. We select all events where two intact protons are observed in the AFP detectors. Both detectors cover the region in ξ such as $0.015 \leq \xi \leq 0.15$ An illustration of this cut is shown in Figure 4.7 representing the histogram the protons ξ for all selected events before and after the *AFP acceptance cut*. The cuts are defined as follow:

```
1 //Defining the AFP cut on the proton xi
2 bool AFP_Cut(double xi_proton){
3 return (AFP_min_xi < xi_proton && xi_proton < AFP_max_xi);}
```

Cuts are applied successively to the dataset; no cuts first, then the AFP cut. The result of the cut is shown in Figure 4.7, where we can see the selection of events with two protons in the AFP acceptance. This cut is not meant to improve the signal-to-noise ratio, but simply reflect the geometrical acceptance of the AFP detector.

```
1 //Applying the cut to both protons
2 vector<bool> cuts;
3 cuts.push_back(true);
```

```
4 cuts.push_back(cuts.back() && Cuts::AFP_Cut(xi1) && Cuts::AFP_Cut(xi2)↩
    );
```

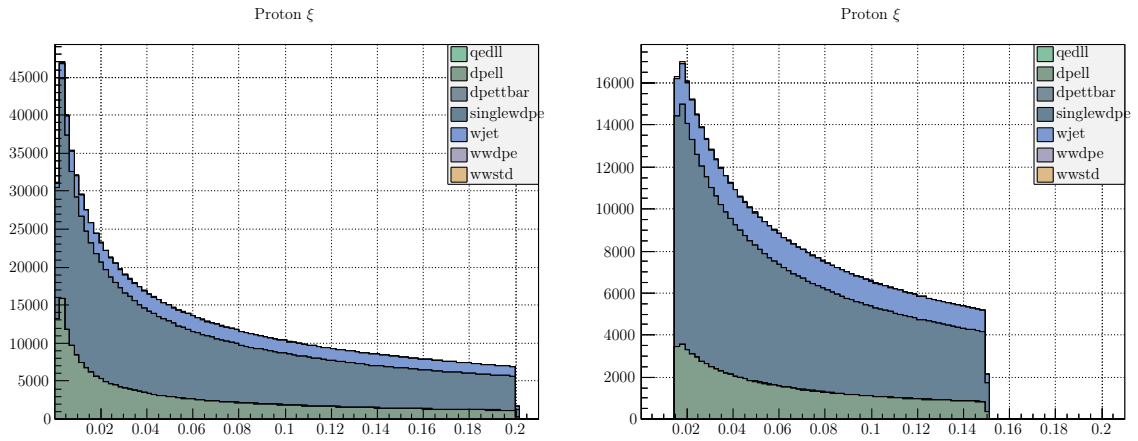


Figure 4.7 – Protons ξ for all selected events before (left) and after (right) the *AFP acceptance cut*. The cut is the expression of the AFP detector acceptance.

4.5.3 Cut on the number of lepton and jets

The second cut selects the semi leptonic case. One W decays into a lepton and the other decays into a $q\overline{q}$ pair, producing two jets. In reality, due to poor reconstruction 2, or more jets can be observed, as well as one or more leptons. The cut is defined as follow:

```
1 //Defining the cut on the number of leptons
2 bool Lepton_Cut(double nb_electrons, double nb_muons){
3   return ((nb_electrons+nb_muons) >= 1);}
4 //Defining the cut on the number of jets
5 bool Jet_Cut(double nb_jet){
6   return (nb_jet >= 2);}
```

This cut is applied to the dataset after the AFP cut. The result of this cut is shown in Figure 4.8 and Figure 4.9, where we can see that events without lepton or a minimum of 2 jets are rejected.

```
1 //Applying the cut on the number of lepton and jets
2 cuts.push_back(cuts.back() && Cuts::AFP_Cut(xi1) && Cuts::AFP_Cut(xi2)↩
    );
3 cuts.push_back(cuts.back() && Cuts::Lepton_Cut(Electrons_, Muons_) && ↩
    Cuts::Jet_Cut(Jets_));
```

This selection cut for semi-leptonic decays reduces the signal by at least 45.6% (Hadronic) + 10.5% (Leptonic) = $\sim$ 66%. The background signals that are significantly suppressed are:

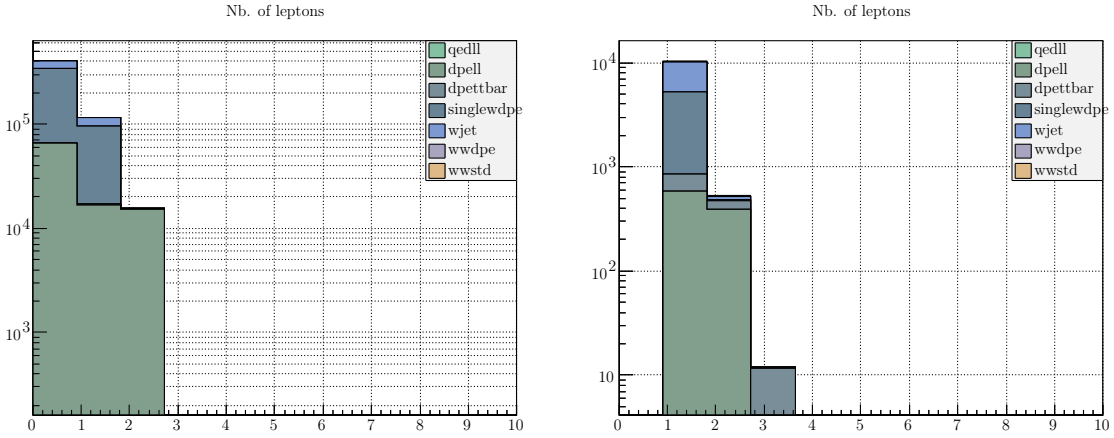


Figure 4.8 – Number of leptons before (left) and after (right) the cut on the number of leptons: $1 \leq lep$.

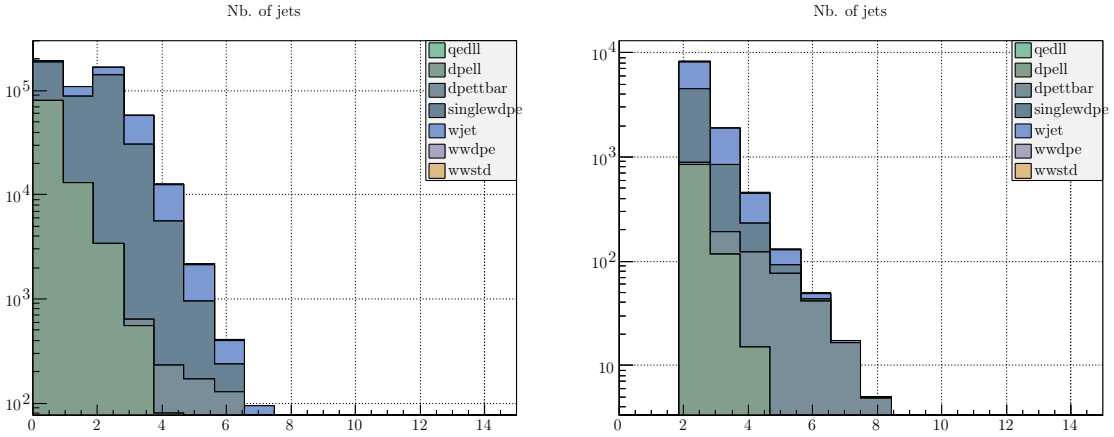


Figure 4.9 – Number of jets before (left) and after (right) the cut on the number of jets: $2 \leq jets$. We note here that the important $\gamma\gamma \rightarrow l\bar{l}$ background for $0jets$ is heavily suppressed by requesting one hadronic decay for the WW pair.

- $DPE \rightarrow l\bar{l}$, the request of two or more jets on the pure leptonic sample leaves only a small fraction of poorly reconstructed event.
- $DPE \rightarrow W + X$, the request of two or more jets and one or more lepton, signature of a WW pair reduces dramatically the background from single W bosons.
- $W + jet$, the request of one or more lepton on the $W + jet$ sample leaves only 10.5% background from the leptonic decays of the W.

4.5.3.1 Cut on p_T of leading lepton or leading jet

As can be seen in Figure 4.10, the leading lepton or jet is generated at higher p_T for the signal than for the $W + jet$ and the $DPE \rightarrow W + X$ backgrounds. Therefore, a

Table 4.2 – Rejection factor for the $1 \leq lep\&2 \leq jets$ cut.

Process	$1 \leq lep\&2 \leq jets$ cut efficiency
$\gamma\gamma \to WW$	78.01%
$DPE \to l\bar{l}$	98.99%
$DPE \to W + X$	98.75%
$W + jet$	94.51%

cut is set at high p_T for either the leading jet or the lepton at the point where the p_T of the signal crosses the background to reduce the $W + jet$ background. That is $p_T^{jet1} \geq 150GeV$ or $p_T^{lep1} \geq 250GeV$.

```
//Defining the cut on the pt of the leading jet
bool Jet_PT_Cut(double jet_pt){
  return (jet_pt >= 150.);}
//Defining the cut on the pt of the leading lepton
bool Lep_PT_Cut(double lep_pt){
  return (lep_pt >= 250.);}
```

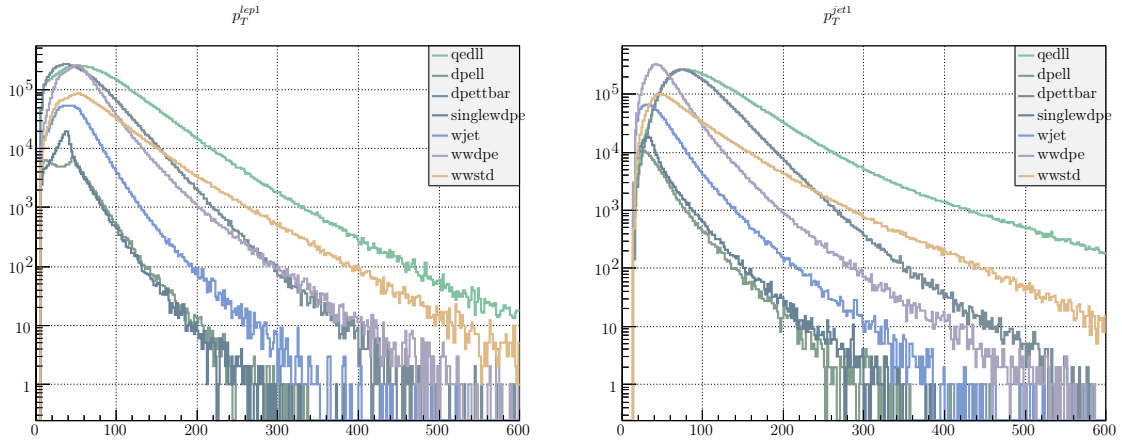


Figure 4.10 – Cumulated plot of the p_T of the leading jet and leading lepton for signal and backgrounds. In both cases the signal has a higher p_T lepton or jet.

This cut is applied after the cut on the number of jets and leptons. The result of this cut is shown in Figure 4.11, where we can see that in either the case of leading jet or the lepton, samples with low p_T jet or lepton are suppressed.

```
//Applying the cut on the pt of the leading jet or lepton
cuts.push_back(cuts.back() && Cuts::Lepton_Cut(Electrons_, Muons_) && ↩
    Cuts::Jet_Cut(Jets_));
cuts.push_back(cuts.back() && (Cuts::Lep_PT_Cut(lep_pt)  || Cuts::↩
    Jet_PT_Cut(Jets_m_PT[jet1])));
```

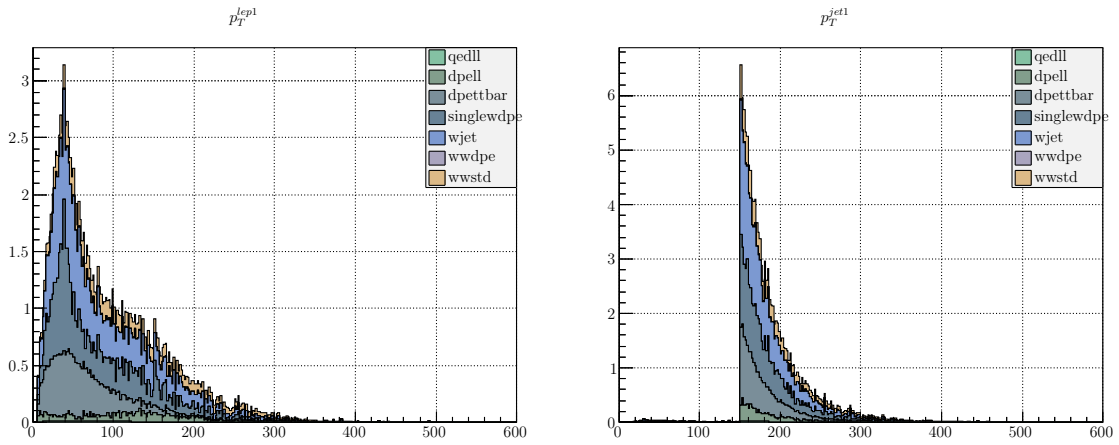


Figure 4.11 – P_T of the leading lepton (left) and jet (right) after the cut $p_T^{lep1} \geq 250GeV \mid p_T^{jet1} \geq 150GeV$.

Although this cut selecting a high p_T jet or lepton could be perceived as very stringent on the signal, it selects a very clean signature of WW pairs created by photon exchange. Backgrounds without a clear high p_T hadron or lepton are heavily suppressed:

Table 4.3 – Rejection factor for the $p_T^{lep1} \geq 250GeV \mid p_T^{jet1} \geq 150GeV$ cut.

Process	$p_T^{lep1} \geq 250GeV \mid p_T^{jet1} \geq 150GeV$ cut efficiency
$\gamma\gamma \to WW$	90.44%
$DPE \to W + X$	99.32%
$W + jet$	99.11%
$DPE \to t\bar{t}$	93.52%
$DPE \to WW$	98.90%
$DPE \to l\bar{l}$	99.34%

4.5.4 Cut on the dijet mass

The dijet system created in the central detector is the decay product of $WW \to q\bar{q} \to dijet$. Therefore reconstruction of the dijet system mass should be within the range of the W rest mass $\sim 80GeV$. In our analysis, the cut on the dijet mass is set at $80 \pm 40GeV$. This cut is applied to the two leading jets.

```
1 //Defining the cut on the dijet mass.
2 bool Dijet_Mass_Cut (double dijet_mass) {
3   return (40. < dijet_mass && dijet_mass < 120.);}
```

The effect of the cut on the dijet mass is shown in Figure 4.12.

```
//Applying the cut on the dijet mass
cuts.push_back(cuts.back() && (Cuts::Lep_PT_Cut(lep_pt)  || Cuts::↩
    Jet_PT_Cut(Jets_m_PT[jet1])));
cuts.push_back(cuts.back() && Cuts::Dijet_Mass_Cut(inv_m_j));
```

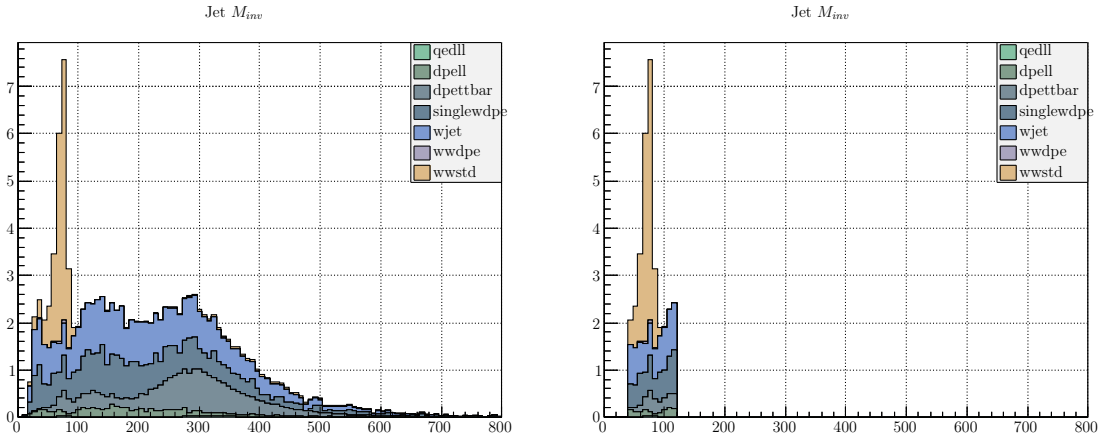


Figure 4.12 – Reconstructed dijet mass before (left) and after (right) the cut $40GeV \leq dijetmass \leq 120GeV$. The cut selects the sharp dijet mass peak around the W mass for the signal (in orange) while removing background events having a dijet mass spread out.

The emission of a WW pair being very costly in energy, produces low momentum W's. Therefore, with an energy close to their rest mass. This not the case for single W production, where the emission is much more boosted resulting in a much flatter spectrum for the dijet mass. The background samples corresponding to the emission of single W are heavily suppressed:

Table 4.4 – Rejection factor for the $40GeV \leq dijetmass \leq 120GeV$ cut.

Process	$40GeV \leq dijetmass \leq 120GeV$ cut efficiency
$\gamma\gamma \to WW$	16.24%
$DPE \to W + X$	80.20%
$W + jet$	81.42%
$DPE \to t\bar{t}$	89.54%
$DPE \to WW$	55.19%
$DPE \to l\bar{l}$	80.66%

4.5.5 Analysis summary

The summary of all cuts on signal and backgrounds is reported in Table 4.5 and Table 4.6. In the analysis we show that ignoring the pile-up background, the signal from WW pair production by photon exchange could be observed with a good sensitivity in the semi-leptonic channel.

It is important to note that this study was conducted without considering pile-up events in the Forward Detectors. At high luminosity (high number of simultaneous interactions, $50 \leq \mu \leq 100$) the background from pile-up events is supposed to be the dominant source of noise. However, we think it will be possible to use the technique used in the leptonic case, namely using the number of tracks outside the two leading jets.

Table 4.5 – Number of signal and background events produced and after each cut.

Cut/process	$\gamma\gamma \to l\bar{l}$	$DPE \to l\bar{l}$	$DPE \to t\bar{t}$	$DPE \to W$	$W + jet$	$DPE \to WW$	$\gamma\gamma \to WW$
production	38.3×10^6	49.8×10^6	42.5×10^6	49.5×10^6	49.5×10^6	49.3×10^6	49.3×10^6
$0.015 \leq \xi \leq 0.15$	23.4×10^6	18.3×10^6	17.1×10^6	21.6×10^6	25.4×10^6	26.9×10^6	15.3×10^6
$1 \leq lep\&2 \leq jets$	12.8×10^6	18.3×10^4	93.7×10^5	27.1×10^4	14.0×10^5	69.2×10^5	33.6×10^5
$p_T^{lep1} \geq 250GeV \mid p_T^{jet1} \geq 150GeV$	24.1×10^5	1215	60.7×10^4	1919	12.3×10^3	76.3×10^3	32.2×10^4
$40GeV \leq dijetmass \leq 120GeV$	41.5×10^4	235	397	380	2288	34.2×10^3	27.0×10^4

Table 4.6 – Number of events for $40fb^{-1}$ after each cut.

Cut/process	$\gamma\gamma \to l\bar{l}$	$DPE \to l\bar{l}$	$DPE \to t\bar{t}$	$DPE \to W$	$W + jet$	$DPE \to WW$	$\gamma\gamma \to WW$
production	9.79	26.5×10^4	1618	81.5×10^4	17.0×10^4	236.1	2748
$0.015 \leq \xi \leq 0.15$	5.98	97.5×10^3	653.1	35.6×10^4	87.6×10^3	128.7	852
$1 \leq lep\&2 \leq jets$	3.26	976	356.5	4.45×10^3	4.81×10^3	33.1	187.5
$p_T^{lep1} \geq 250GeV \mid p_T^{jet1} \geq 150GeV$	0.62	6.48	23.1	31.6	42.36	0.37	17.9
$40GeV \leq dijetmass \leq 120GeV$	0.11	1.25	2.41	6.25	7.87	0.16	15.0

4.6 Beyond the standard model

4.6.1 Theoretical context

Good signal to noise observation in the semi-leptonic channel encourages us to look at the potential new physics. It is possible to modify the trilinear and quartic gauge couplings introduced in Equation 4.4 and Equation 4.5 to simulate the effect at low energy of an unknown theory *Beyond Standard Model* (BSM). The lowest order interaction Lagrangians involving two photons are dimension-6 operators [6]:

$$\mathcal{L}_6^0 = \frac{-e^2}{8}\frac{a_0^W}{\Lambda^2}F_{\mu\nu}F^{\mu\nu}W^{+\alpha}W^-_\alpha - \frac{e^2}{16\cos^2\theta_W}\frac{a_0^Z}{\Lambda^2}F_{\mu\nu}F^{\mu\nu}Z^\alpha Z_\alpha$$
$$\mathcal{L}_6^C = \frac{-e^2}{16}\frac{a_C^W}{\Lambda^2}F_{\mu\alpha}F^{\mu\beta}(W^{+\alpha}W^-_\beta + W^{-\alpha}W^+_\beta) - \frac{e^2}{16\cos^2\theta_W}\frac{a_C^Z}{\Lambda^2}F_{\mu\alpha}F^{\mu\beta}Z^\alpha Z_\beta$$

where a_0, a_C are the parametrized, new coupling constants. A new scale Λ is introduced so that the Lagrangian density has the correct dimension four and is interpreted as the typical mass scale of new physics. Note only quartic anomalous couplings are considered in this study. In the above formula, we assigned the W and Z parts of the Lagrangian to have specific couplings, i.e. $a_0 \to (a_0^W, a_0^Z)$ and similarly $a_C \to (a_C^W, a_C^Z)$.

The WW and ZZ two-photon cross sections rise quickly at high energies when any of the anomalous parameters are non-zero [10]. The cross section rise has to be regulated by a form factor, which vanishes in the high energy limit to construct a realistic physical model of the BSM theory. We therefore modify the couplings by form factors that have the desired behavior, i.e. they modify the coupling at small energies only slightly but suppress it when the center-of-mass energy $W_{\gamma\gamma}$ increases. The form of the form factor that we consider is the following

$$a \to \frac{a}{(1 + W_{\gamma\gamma}^2/\Lambda^2)^n} \tag{4.7}$$

there are no preferred values for n and Λ, in the following we will use: n=2, and Λ ~2 TeV.

4.6.2 Anomalous Monte-Carlo samples

Similar to Standard Model WW production, the anomalous coupling signal is generated using FPMC. Anomalous coupling terms are set in the FPMC datacards. We have considered the four anomalous coupling values: $a_0^W = 5.10^{-6}, 10^{-6}, 5.10^{-7}, 10^{-7}$. The same preselection on the samples is used: $PTMIN = 20GeV$ and $YWWMIN = 0.001$ and $YWWMAX = 0.2$. The background in this analysis is the same as before, with the Standard Model now considered as a background as well. For each signal and background 50 M events have been generated and reconstructed.

4.6.3 Anomalous signal selection

The cut flow for the anomalous signal selection is similar to the one used for the standard signal selection:

- Cut 1: $0.0015 \leq \xi \leq 0.15$.
- Cut 2: $1 \leq lep\&2 \leq jets$.
- Cut 3: $p_T^{lep1} \geq 250GeV \mid p_T^{jet1} \geq 150GeV$.
- Cut 4: $40GeV \leq dijetmass \leq 120GeV$.

At this point we are left almost exclusively with the standard and the anomalous signal events. A last cut on the invariant mass M_X is added, taking advantage of the high mass of the anomalous signal.

- Cut 5: $M_X \geq 700GeV$.

The result of the successive cuts on the signal and background for an anomalous coupling value $a_0^W = 5.10^{-6}$ are shown in Figure 4.16. Tables 4.7 and 4.8 tabulate the number of events after each cut for signal and background. Table 4.8 reports the number of events for an integrated luminosity of $\mathcal{L} = 40fb^{-1}$. At $a_0^W = 5.10^{-6}$, we are left with 37 semi-leptonic events and 3 background events (from Standard Model). This corresponds to a 5 σ observation, similar to the sensitivity obtained in the leptonic case.

We have to *remind here* that this study was done without the contribution of pile-up events in the background. A full simulation including pile-up events, should be carried out in order to confirm that this sensitivity can indeed be obtained in ATLAS with the AFP detector.

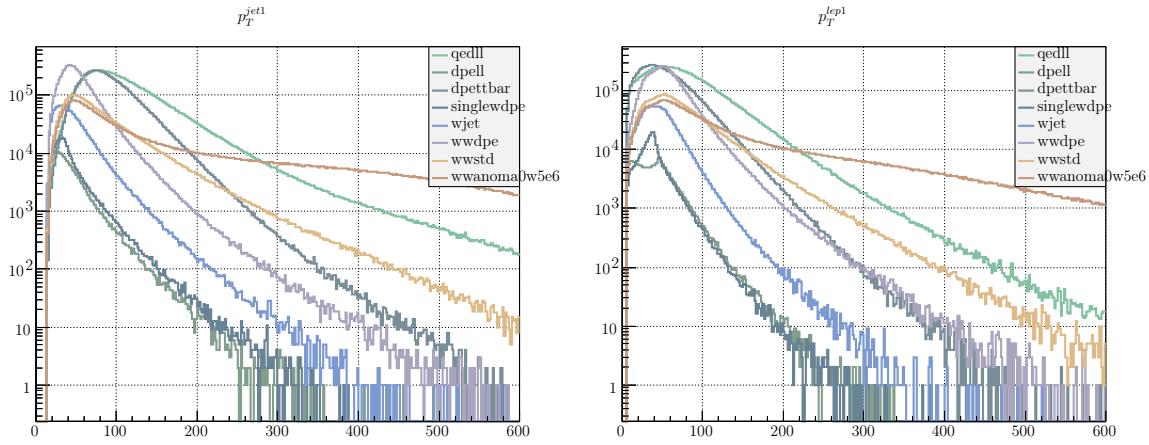


Figure 4.13 – Stacked plot of the p_T of the leading jet and leading lepton for all signal and background after the AFP cut. In both cases the signal has a high p_T lepton or jet. We can note already that the anomalous coupling signal at $a_0^W = 10^{-7}$ follows the standard one.

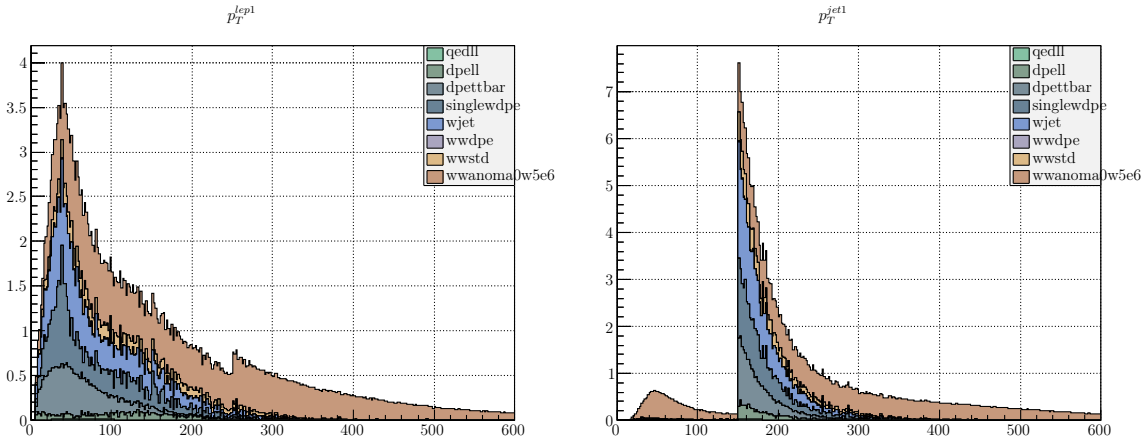


Figure 4.14 – P_T of the leading lepton (left) and jet (right) after the cut $p_T^{lep1} \geq 250GeV$ | $p_T^{jet1} \geq 150GeV$.

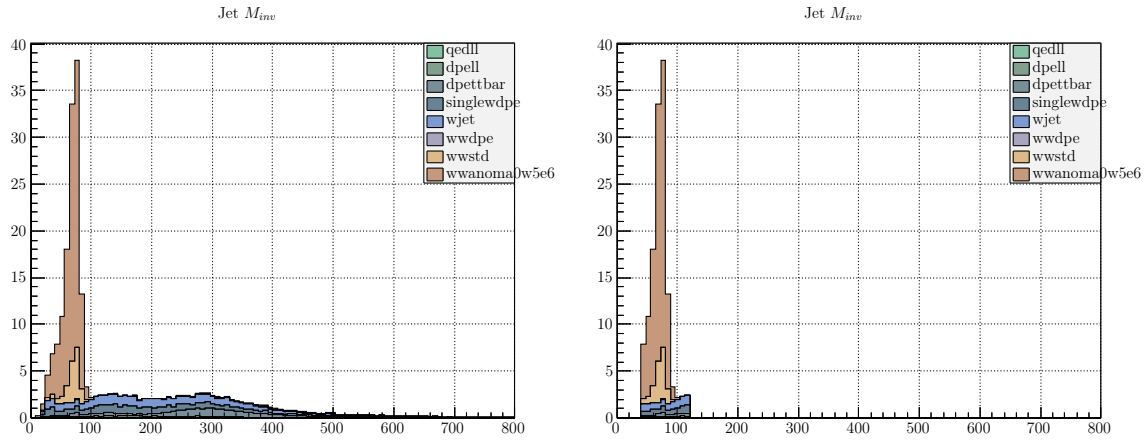


Figure 4.15 – Plot of the reconstructed dijet mass before (left) and after (right) the cut $40GeV \leq dijetmass \leq 120GeV$.

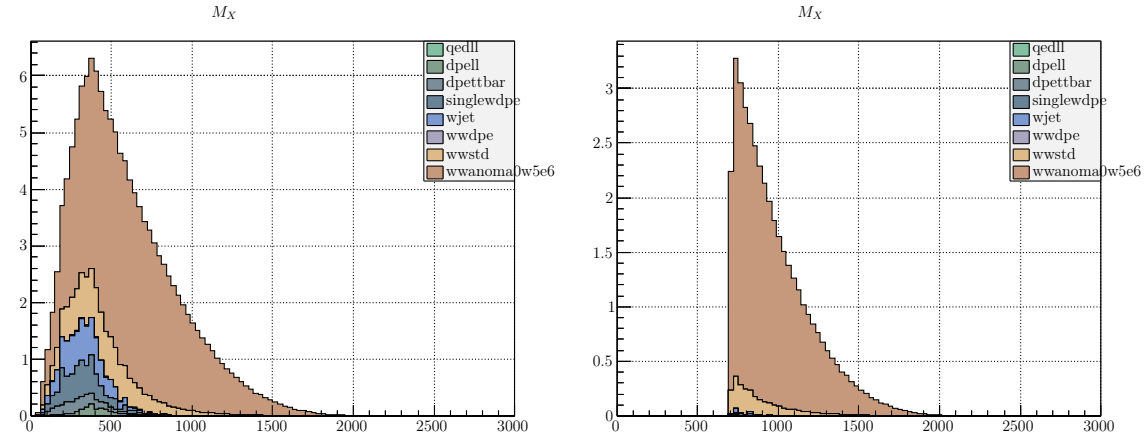


Figure 4.16 – Plot of the mass of central object M_X (dijet + lepton) before (left) and (after) the cut $M_X \geq 700GeV$.

Table 4.7 – Number of signal and background events produced and after each cut, for an anomalous coupling value $a_0^W = 5.10^{-6}$.

Cut/process	$\gamma\gamma \to l\bar{l}$	$DPE \to l\bar{l}$	$DPE \to t\bar{t}$	$DPE \to W$	$W + jet$	$DPE \to WW$	$\gamma\gamma \to WW$	WW anom
production	38.3×10^6	49.8×10^6	42.5×10^6	49.5×10^6	49.5×10^6	49.3×10^6	49.3×10^6	49.3×10^6
$0.0015 \leq \xi \leq 0.15$	23.4×10^6	18.3×10^6	17.1×10^6	21.6×10^6	25.4×10^6	26.9×10^6	15.3×10^6	18.0×10^6
$1 \leq lep\&2 \leq jets$	12.8×10^6	18.3×10^4	93.7×10^5	27.1×10^4	14.0×10^5	69.2×10^5	33.6×10^5	39.8×10^5
p_T *cut*	24.1×10^5	1215	60.7×10^4	1919	12.3×10^3	76.3×10^3	32.2×10^4	16.9×10^5
dijet mass cut	41.5×10^4	235	63.5×10^3	380	2288	34.2×10^3	27.0×10^4	13.5×10^5
$M_X \geq 700GeV$	24.1×10^3	3	397	3	31	1284	44.3×10^3	51.4×10^4

Table 4.8 – Number of events for $40fb^{-1}$ after each cut for an anomalous coupling value $a_0^W = 5.10^{-6}$.

Cut/process	$\gamma\gamma \to l\bar{l}$	$DPE \to l\bar{l}$	$DPE \to t\bar{t}$	$DPE \to W$	$W + jet$	$DPE \to WW$	$\gamma\gamma \to WW$	WW anom
production	9.79	26.5×10^4	1618	81.5×10^4	17.0×10^4	236.1	2748	3608
$0.0015 \leq \xi \leq 0.15$	5.98	97.5×10^3	653.1	35.6×10^4	87.6×10^3	128.7	852	1315
$1 \leq lep\&2 \leq jets$	3.26	976	356.5	4.45×10^3	4.81×10^3	33.1	187.5	291
p_T *cut*	0.62	6.48	23.1	31.6	42.36	0.37	17.9	124
dijet mass cut	0.11	1.25	2.41	6.25	7.87	0.16	15.0	99.0
$M_X \geq 700GeV$	0.01	0.02	0.02	0.05	0.11	0.01	2.47	37.6

4.6.4 Anomalous coupling reach

As mentioned before, production of the anomalous coupling samples has been done for the values: $a_0^W = 5.10^{-6}, 10^{-6}, 5.10^{-7}, 10^{-7}$. For couplings below 5.10^{-7}; the anomalous coupling signals becomes increasingly similar to the standard model, and the high invariant mass cut in particular becomes no longer effective.

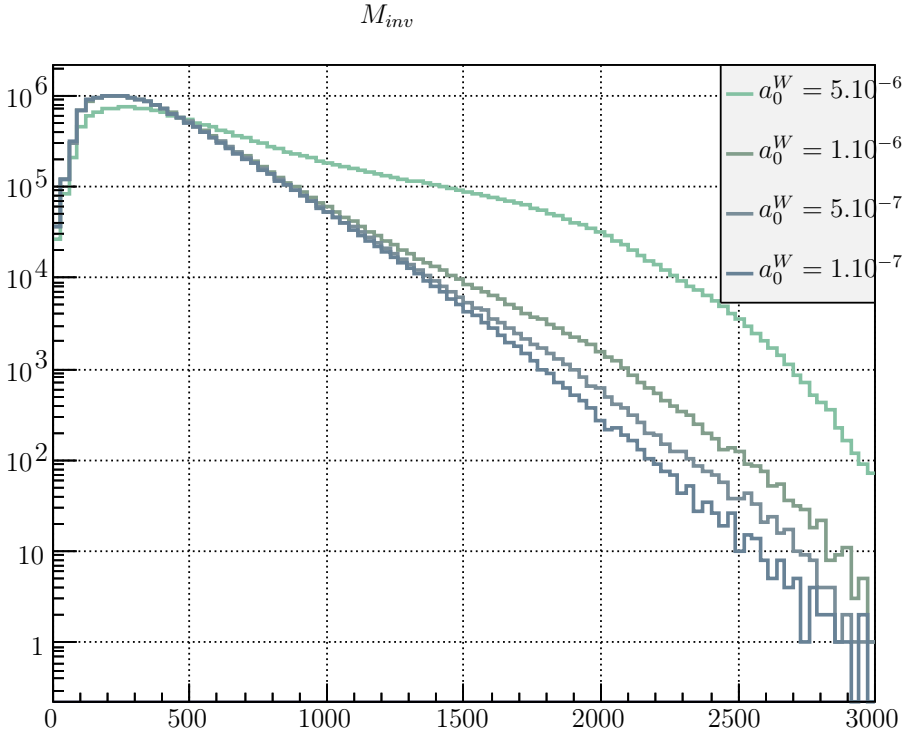


Figure 4.17 – Invariant mass M_{inv} of the anomalous coupling samples for the values: $a_0^W = 5.10^{-6}, 10^{-6}, 5.10^{-7}, 10^{-7}$. For couplings below 5.10^{-7}; the anomalous coupling signals becomes increasingly similar to the standard model, and the high invariant mass cut in particular becomes no longer effective.

4.7 Conclusion

The study of the semi-leptonic decay of the WW pair production by photon exchange: $pp \to p\gamma\gamma p \to pW^+W^-p$, has been done with the FMPC generator and Fast ATLAS reconstruction. The result shows similar performances compared to the leptonics studies (see Table 4.10) in the measurement of anomalous coupling in ATLAS using the AFP detector. Combined with the full leptonic decay mode, it promises to increase the resolution in the measurment of the quartic anomalous coupling term a_0^W.

The full study of this process in ATLAS integrating pile-up events and Full Reconstruction has yet to be done. According to the preliminary results presented here,

Table 4.9 – Number of events for the standard model and anomalous coupling values from $a_0^W = 5.10^{-6}$ to $a_0^W = 10^{-7}$ after each cut. Below $a_0^W = 5.10^{-7}$ no significant gain on the signal to noise is expected using the invariant mass cut.

Cut/process	*std. WW*	$a_0^W = 5.10^{-6}$	$a_0^W = 10^{-6}$	$a_0^W = 5.10^{-7}$	$a_0^W = 10^{-7}$
production	2748	3608	2789	2760	2744
$0.0015 \leq \xi \leq 0.15$	852	1315	874	859	852
$1 \leq lep\&2 \leq jets$	187.5	291	192	189	187.2
p_T *cut*	17.9	124	22.6	19.1	18.0
dijet mass cut	15.0	99.0	18.7	16.0	15.1
$M_X \geq 700GeV$	2.47	37.6	3.88	2.82	2.48

we believe that using a similar approach to the work presented for the leptonic case [1] - using 10-picosecond timing detector to identify primary vertex and fitting the dijet and lepton track to the vertex - a resolution comparable to the one presented could be obtained. See Table and 4.10 showing the cut flow and reach obtained in the leptonic case.

Table 4.10 – Number of expected signal and background events for 300 fb^{-1}, for a pile-up of $\mu = 46$. A time resolution of 10 ps has been assumed for background rejection. The diffractive background comprises production of QED diboson, QED dilepton, diffractive WW, double pomeron exchange WW.

Cuts	Top	Dibosons	Drell-Yan	W/Z+jet	Diffr.	$a_0^W = 5.10^{-6}$
timing $\leq 10ps$						
$p_T^{lep1} \geq 150GeV$	5198	601	20093	1820	190	282
$p_T^{lep2} \geq 20GeV$						
$M(ll) \geq 300GeV$	1650	176	2512	7.7	176	248
nTracks≤ 3	2.8	2.1	78	0	51	71
$\Delta\phi \leq 3.1$	2.5	1.7	29	0	2.5	56
$M_X \geq 800GeV$	0.6	0.4	7.3	0	1.1	50
$p_T^{lep1} \geq 300GeV$	0	0.2	0	0	0.2	35

Bibliography

[1] O Kepka, C Royon, L Schoeffel, R Staszewski, M Trzebinski, and R Zlebcik. *Physics Cases within the AFP project*. Technical Report ATL-COM-PHYS-2012-775, CERN, Geneva, Jun 2012. ATLAS note to be made public.

[2] T Pierzchała and Krzysztof Piotrzkowski. *Sensitivity to anomalous quartic gauge couplings in photon-photon interactions at the LHC*. *Nuclear Physics B-Proceedings Supplements*, 179:257–264, 2008.

[3] M Boonekamp, F Chevallier, C Royon, and L Schoeffel. *Understanding the structure of the proton: From HERA and Tevatron to LHC*. *arXiv preprint arXiv:0902.1678*, 2009.

[4] Laurent Schoeffel. *Advances in diffraction of subnuclear waves*. *Progress in Particle and Nuclear Physics*, 65(1):9–49, 2010.

[5] VM Budnev, IF Ginzburg, GV Meledin, and VG Serbo. *The two-photon particle production mechanism. Physical problems. Applications. Equivalent photon approximation*. *Physics Reports*, 15(4):181–282, 1975.

[6] E Chapon, C Royon, and O Kepka. *Anomalous quartic $WW\gamma\gamma$, $ZZ\gamma\gamma$, and trilinear $WW\gamma$ couplings in two-photon processes at high luminosity at the LHC*. *Physical Review D*, 81(7):74003, 2010.

[7] O Kepka, C Royon, M Saimpert, and R Zlebcik. *Multiple proton-proton interactions in Atlas Forward Detector pp at $\sqrt{s} = 14\,\mathrm{TeV}$*. Technical Report ATL-COM-PHYS-2013-847, CERN, Geneva, Jun 2013.

[8] Oldrich Kepka. *QCD and Diffraction in the ATLAS Experiment at the LHC*. PhD thesis, Orsay, U. PARIS XI, Orsay, 2009.

[9] M Boonekamp, A Dechambre, V Juranek, O Kepka, M Rangel, C Royon, and R Staszewski. *FPMC: a generator for forward physics*. 2011.

[10] Oscar JP Eboli, Ma Concepción Gonzalez-García, SM Lietti, and SF Novaes. *Anomalous quartic gauge boson couplings at hadron colliders*. *Physical Review D*, 63(7):075008, 2001.

Part II

Performance of detectors and algorithms for picosecond time measurements

5

Micro-channel plate photo-detector

Contents

5.1 Introduction

In this Chapter we will describe one kind of promising photo-detector for picosecond time-of-flight measurment: the micro-channel plate photo-detector. This detector converts promptly photon into a cloud of 10^5 to 10^7 electrons, measurable by external electronics. As we will see, the fast process guarantees a low time jitter between the time of the photon incidence on the detector and the electrical signal measurement, providing a precise time measurement.

This Chapter is divided in eight Sections. Section 5.2 presents the basics of photo-detectors, Section 5.3 the micro-channel plate detector structure, and its different parts: the glass window Section 5.4, the photocathode Section 5.5, the micro-channel plate Section 5.6. Section 5.7 and 5.8 reports the time properties of micro-channel plates detectors and their anodes Section 5.9.

5.2 Description of photo-detectors

5.2.1 Introduction

A photo-detector is a device converting photons into a detectable signal [1]. The signal can be a change of temperature in *bolometers*, a change of resistance in *photo-resistances* or more commonly an electrical signal. The creation of the electric signal from a photon relies on two types of effects:

- **The photoelectric effect** A photon hitting a metal with more energy than the binding energy of an electron to the metal will eject it. Effectively converting the uncharged photon signal into a charged signal: an electron.

- **The electron-holes pair creation** A photon hitting a semi-conductor with more energy than its band-gap energy (energy required to go from the valence band to the conduction band) will create mobile carriers (electrons and holes) effectively creating a charged signal.

In the case of the photoelectric effect, the electric signal is small: equal to the charge of the created photo-electrons. In most cases, amplification is required to get a measurable signal.

In the case of electron-hole pair creation, one photon, given enough energy can generate enough electron-hole pairs for a direct measurement, such as in the Charge Coupled Device (CCD) image sensors.

However in most cases (photoelectric or via the electron-hole creation), the number of generated electrons is not sufficient for a good measurement and amplification is required. Two different amplification techniques can be used:

- **Secondary emission** The amplification process relies on the principle of secondary emission: the photo-electron or the electron of the electron-hole pair produced is accelerated by an electric field into an emissive material that produces secondary electrons by impact ionization. The produced electron cascade is amplified exponentially in the process and high gains are achieved. This principle is illustrated for a micro pore in Figure 5.1.

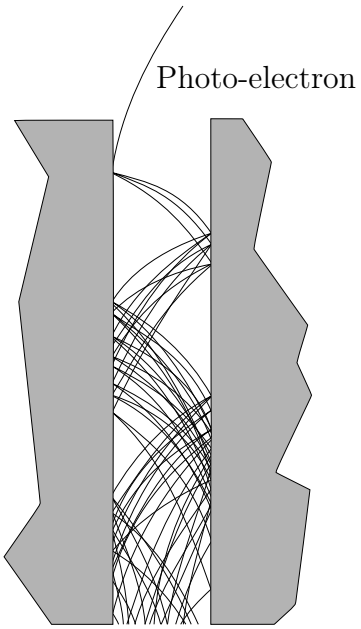


Figure 5.1 – Illustration of the principle of secondary emission amplification of electrons in a micro-channel plate pore. A photoelectron is accelerated towards the wall of the pore and will produce secondaries upon impact. An electrical field accelerates the electrons produced and the process repeats itself increasing exponentially the number of electrons produced along the pore.

- **Electronic amplifier** This amplification process relies on a external device increasing the power of the signal produced by the photo-detector. The amplifier

solution however will amplify both the signal and the noise at its input, and will not improve the signal to noise ratio, if no additional filtering is applied.

One example of a photo-detector is the Micro-Channel Plate Photo-Multiplier (MCP-PMT) which has the advantage of high gain and very good transit time resolution and will be described in the following Sections.

5.3 Micro-channel plate photo-detector

A micro-channel plate photo-detector is illustrated in Figure 5.2. It consists of four different parts: an input window, a photo-cathode, a pair of micro-channel plates and an anode. All part are described in detail in subsequent Sections.

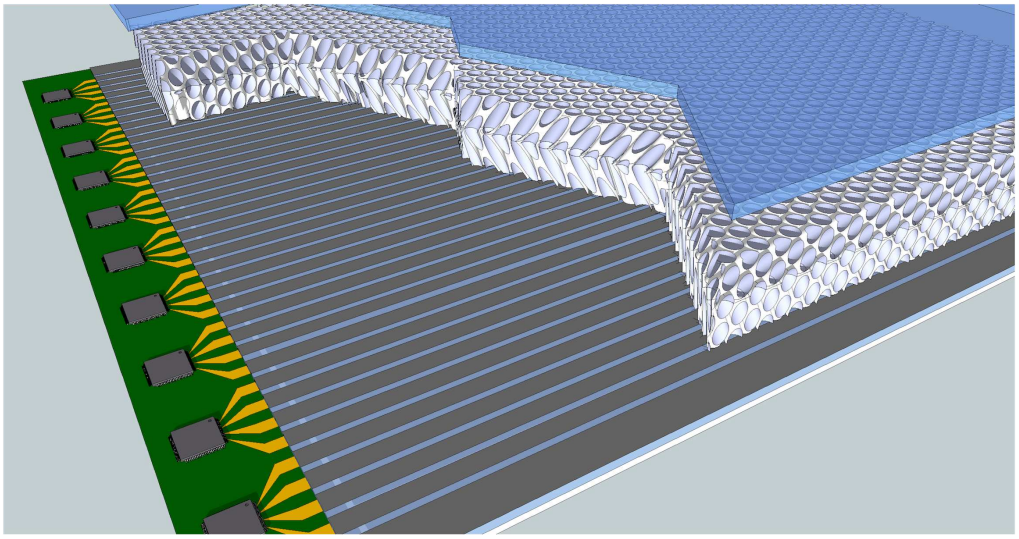


Figure 5.2 – Conceptual sketch of a large area micro-channel plate detector, consisting of an input window with photocathode deposited on the inside, a stack of micro-channel plates and a micro-strip line array connected to the readout electronics.

Due to the ionization processes involved in the micro-channels (see Figure 5.1), all the active parts of the detector must reside in vacuum. Therefore a transparent *input window* is required, separating the active parts of the detector from the outside world, while letting the photons of interest go inside the detector.

The *photo-cathode* is the material, which converts the photon into electrons (also called photo-electrons). *Micro-channel plates* contain the micro-pores where the electrons are amplified by the process of secondary emission. In the end, the electrons produced are collected in the *anodes*.

5.3.1 The LAPPD micro-channel plate photo-detector

The micro-channel plate photo-detector that will be presented in this Section is a 20 cm square glass detector developed by the Large-Area Picosecond Photo-Detector (LAPPD) collaboration. This collaboration consists of physicists and engineers in three US national laboratories; Argonne, Fermilab and SLAC, five universities: UC Berkeley, Chicago, Illinois at Chicago, Illinois Champaign-Urbana, Hawaii and Washington at St. Louis as well as three small US companies: Arradiance, Muons Inc and Synkera [2] [3].

The first year of the thesis was spent at the University of Chicago, as a member of the LAPPD collaboration, where the anode part of the detector (see Section 5.9) was studied and its bandwidth improved [4] [5]. An important contribution was to the readout electronics envisioned to equip the detector: the PSEC ASIC family (see Chapter 8) [6], [7], [8].

The design of the LAPPD micro-channel plate photo-detector is strongly motivated by developing large-area photo-detectors with excellent time resolution (few picoseconds) [9], [10]. The ATLAS Forward Physics project (see Chapter 3) would strongly benefit from the developments of the LAPPD collaboration, as this kind of detector could be used in association with the quartz radiators in the time-of-flight detector.

A photograph of the detector without photocathode or micro-channel plates, is showed in Figure 5.3

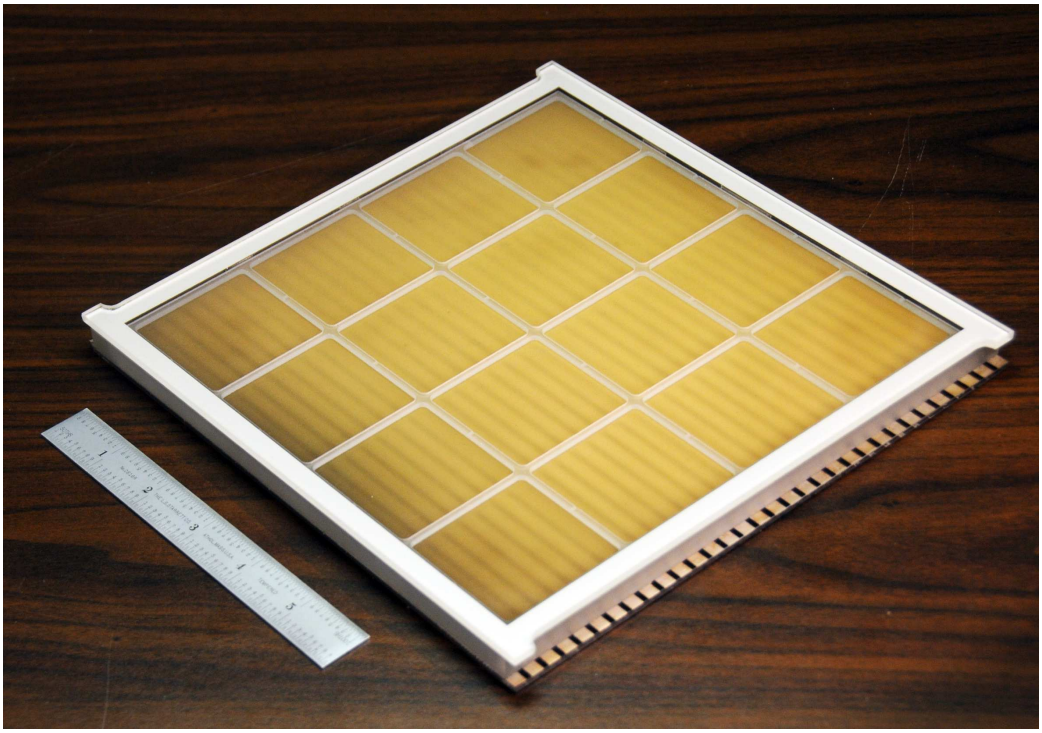


Figure 5.3 – Photograph of the 8 inch LAPPD photo-detector, without photo-cathode deposited or micro-channel plates installed.

5.4 Glass window

The process of photo-electron emission and electron amplification must take place in vacuum, therefore the photo-detector must be encased, with a transparent window to let the photon of interest through.

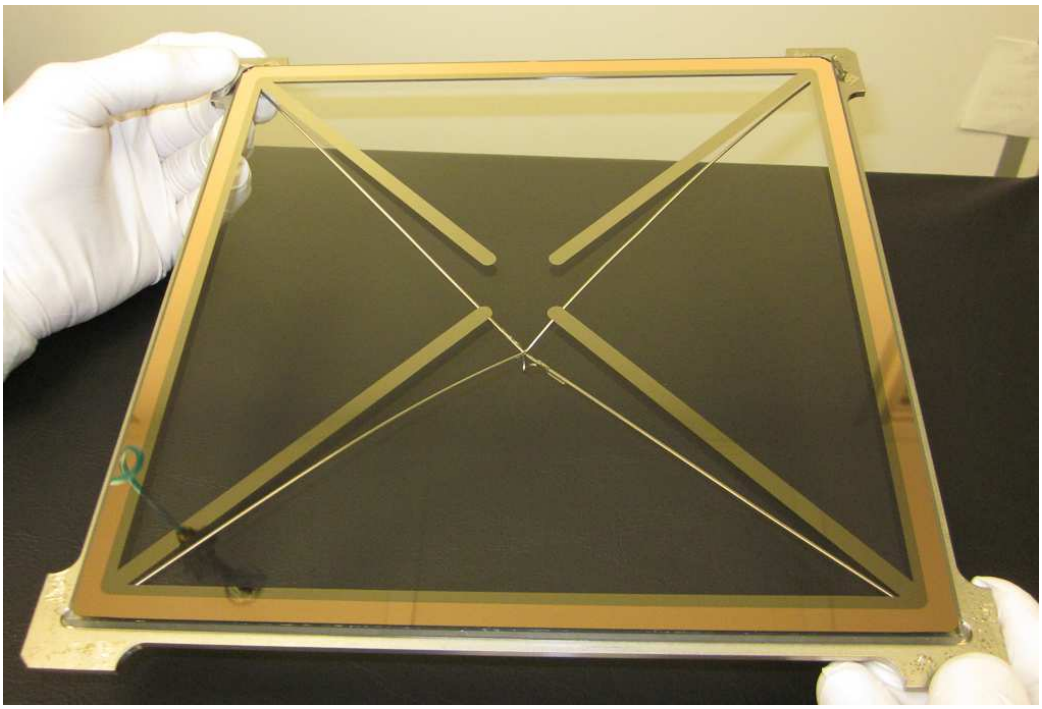


Figure 5.4 – Photograph of an 8 inch square Borofloat 33 from Schott window with electrode deposited for high voltage distribution to the photocathode.

5.4.1 Window attenuation and cutoff frequency

The window must be made of transparent material at the wavelength of the photon of interest. For example, in Figure 5.5 is plotted the transmission efficiency through the NF-K5 glass from Schott [11]. This Figure illustrates a typical tradeoff in photo-detector design, where a thicker input window will be more absorbant than thinner versions, for all wavelengths. However, since the inside of the photo-detector is in vacuum, a high pressure is exerted on the window of the detector and a thick window is required to avoid any cracks or breaks when operating in atmospheric pressure conditions.

The input window and the photo-cathode are often a monolithic component: a window of quartz or glass with a deposited or evaporated photo-cathode.

Typical materials used for the window are:

- **MgF_2 crystal** Crystals of alkali halide are superior in transmitting ultraviolet radiation, but have the disadvantage of hygroscopy (sensitivity to water). Magnesium Fluoride (MgF_2) is usually used because it shows very low hygroscopy and allows transmission of ultraviolet radiation down to 115 nm.

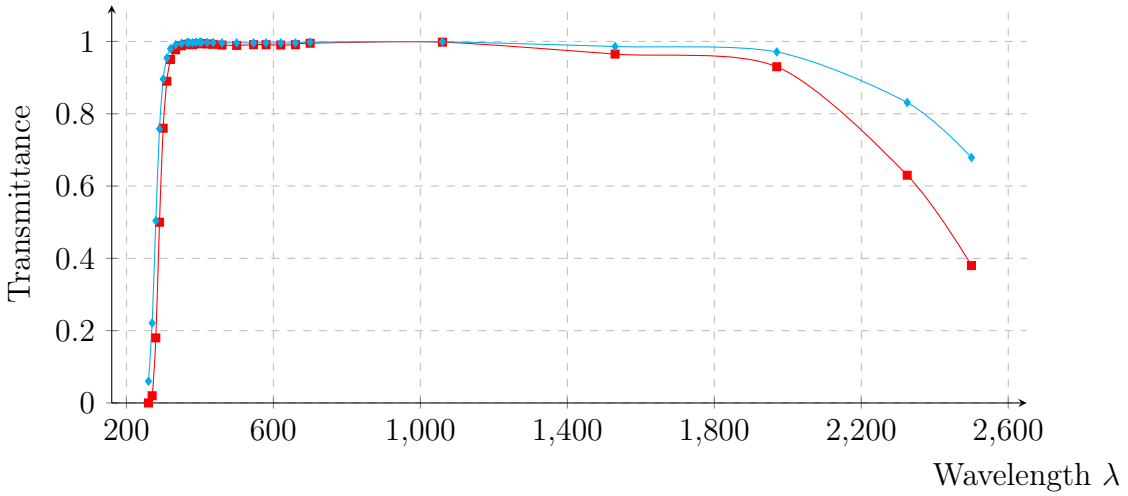


Figure 5.5 – Transmission of the NF-K5 glass from Schott [11] for a thickness of 10mm (—♦—) and 25mm (—■—).

- **Sapphire** Sapphire is made of Al_2O_3 crystal with a lower transmittance, bounded at 150 nm. Because of its extreme surface hardness, high thermal conductivity sapphire is ideal for demanding applications.

- **Synthetic silica** (or fused quartz), transmits ultraviolet light down to 160 nm. Silica is usually not suitable for assembly with the body of the photo-detector as it has a thermal expansion coefficient greatly different from the Kovar metal alloy used for pins and high voltage connections.

- **UV-transmitting glass** Borosilicate glasses specially designed to be transparent to ultraviolet radiation, e.g. Schott 8337. The ultraviolet cut-off frequency is usually around 180 nm.

- **Borosilicate glass** The most frequently used material, because it has the same thermal coefficient as the Kovar alloy used for tube stems (pins). It transmits light from infrared down to approximately 300 nm. For some applications where radioactivity of potassium 40 (K40) contained in the glass affects the measurement, "K-free" borosilicate must be used.

The different cutoff frequencies in the ultraviolet domain of the different types of glass are plotted in Figure 5.6.

In conclusion, the glass window must be chosen in order to be able to measure the wavelengths of interest. If the typical wavelenght of the photon of interest falls in the 200 nm domain, a borosilicate glass will be excluded, as it will be completely opaque to these photons.

In the LAPPD project, the glass chosen was a borosilicate glass, the Borofloat33 from Schott [11], which has the advantage of a similar thermal expansion to the kovar alloy and be relatively low price.

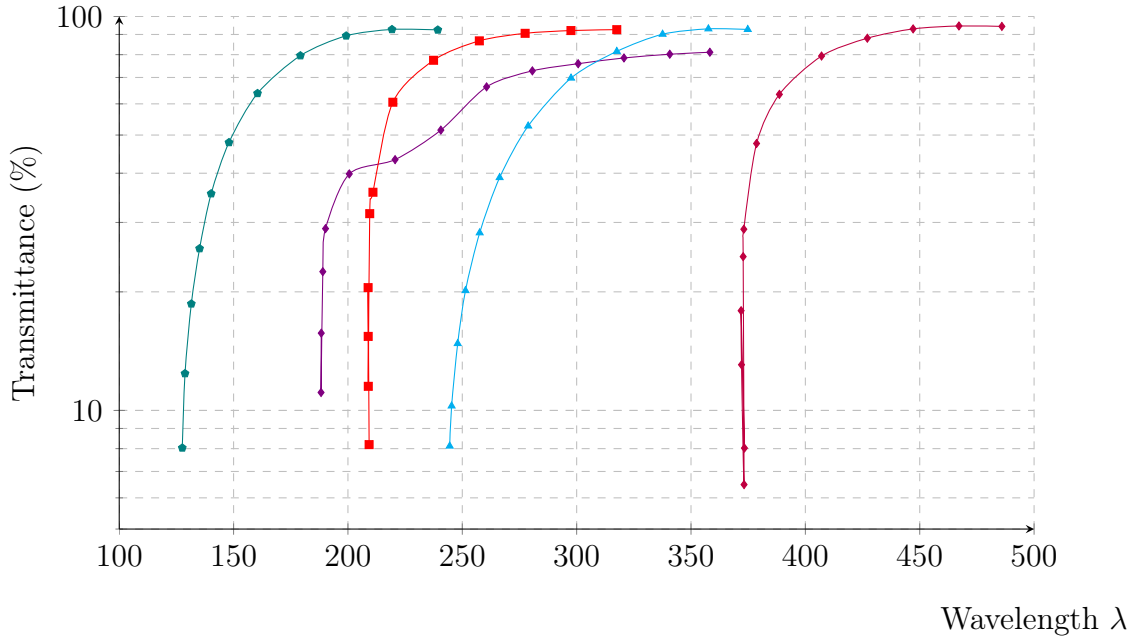


Figure 5.6 – Spectral transmittance of window materials versus the wavelength λ for MgF_2 (—●—), Synthetic silica (—♦—), Sapphire (—■—), UV-Transmitting glass (—▲—) and Borosilicate (—♦—).

5.4.2 Thermal expansion and enclosing

Micro-channel plates photo-detectors are required to work in vacuum. Therefore in most cases the input window must be sealed to the photo-detector case. Sealing is a very delicate operation during which we are facing several difficulties:

- **Window thermal expansion** The window thermal expansion must be closely matched to the detector body, especially for large area detectors. In case of a mismatch, such as between the synthetic silica and the borosilicate glass, a graded seal using graded thermal expansion coefficient materials can be made. But the graded seal is very fragile and a proper care must be taken when handling the detector.

- **Seal** Sealing is a delicate operation, which must be performed in vacuum without increasing the temperature too much to avoid damaging the photocathode and using a seal material with a thermal coefficient similar to the window and the body of the detector. Solder glasses (or frit glasses) with low softening point, below 550°C, are usually used. They achieve sealing without creating thermal damage and matching closely the coefficient of thermal expansion of both pieces.

- **Detector body** The material making the detector body must be chosenin order to ensure a close match to the input window thermal expansion, while also permitting high voltage and anode connections to the outside world. Borosili-

cate is frequently chosen due to its good match to the Kovar, which is usually used for connections.

For evaluation of the characterization of the micro-channel plates and the anode readout network, it is simpler to have the entire detector sitting in vacuum, thus avoiding the need for a seal. Figure 5.7 illustrates an experimental setup used at the Argonne National Laboratory, where no seal is required: the entire photocathode, micro-channel plate and anode stack are placed in a hermetic box whith external vacuum is applied.

Figure 5.7 – Photograph of a test bench where no seal is required for the micro-channel plate detector. The entire stack of components is placed in a hermetic box with external vacuum applied.

For the Large Area Picosecond Photo-detector project the choice was made to build an integrated detector out of the same, cheap glass material: the borosilicate glass B33 from Schott. A huge advantage of this choice is to avoid completely thermal

expansion problems, which are a big issue especially for large area detectors. A picture of the detector body parts illustrates this design option in Figure 5.8

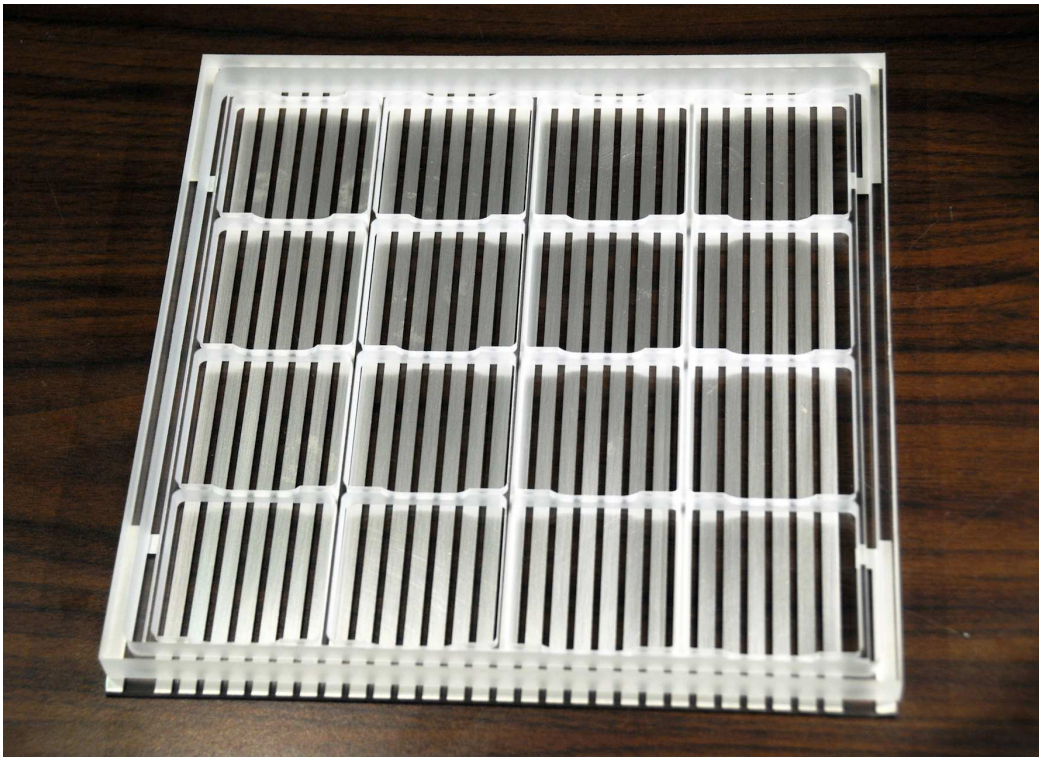


Figure 5.8 – Photograph of the glass components of the 8 inch square micro-channel plate detector, all made of Borofloat33 from Schott.

5.4.3 Conclusion

In summary, the choice to use borosilicate glass (B33 from Schott) was made by the Large Area Picosecond Photo-detector project in order to keep the overall cost of the detector down and to avoid thermal expansion issues, critical for a large area vacuum detectors.

5.5 Photocathode

Single photon signals are too small for direct detection, therefore the photon is converted into an electron, also known as *photoelectron*. This is done using an absorbtion layer called a photocathode. The photocathode is deposited on the inner side of the input window.

5.5.0.1 Spectral response

The photocathode of a micro-channel plate converts the energy of the incident light into photoelectrons using the photoelectric effect. A photograph of a 20 cm square photocathode taken right after deposition is shown in Figure 5.9.

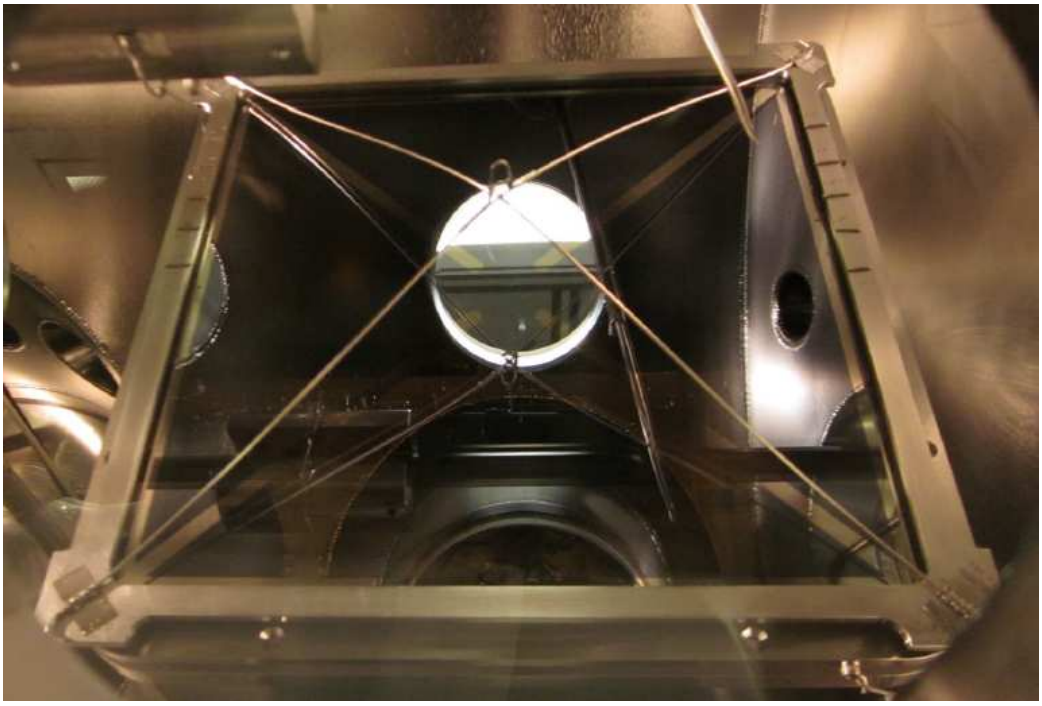


Figure 5.9 – Photograph of a photocathode taken right after deposition on an 20 cm by 20 cm glass window [12].

The conversion efficiency varies with the wavelength of the incident light. The relation between the photocathode sensitivity and the wavelength is called the spectral response. The spectral response is determined in the large wavelength range by the photocathode type and thickness and in the short wavelength range by the input window transmission. Typical spectral sensitivities characteristic of standard photocathodes associated with window materials are shown in Figure 5.10.

5.5.0.2 Radiant sensitivity and quantum efficiency

Spectral response is usually expressed in terms of quantum efficiency and radiant sensitivity. The quantum efficiency is defined as the ratio of the number of photo-

electrons emitted from the photocathode to the number of incident photons. It is usually reported in percentage.

$$\text{QE} = \frac{\text{Number of photoelectrons}}{\text{Number of photons}} \times 100(\%) \tag{5.1}$$

The radiant sensitivity (S) is the photoelectric current from the photocathode divided by the incident radiant power at a given wavelength, expressed in $A.W^{-1}$ (ampere per watt).

$$\text{S} = \frac{\text{Photoelectric Current}}{\text{Radiant Power of Light}}(A.W^{-1}) \tag{5.2}$$

Quantum efficiency and Radiance have the following relationship at a given wavelength λ:

$$\text{QE} = \frac{\text{S} \times 1240}{\lambda} \times 100(\%) \tag{5.3}$$

Typical values for a wide range of photocathode are shown in Figure 5.10.

5.5.1 Photocathode material

Most photocathodes are made of compound semiconductors which consist of alkali metals with a low work function. Approximately ten types of photocathodes are used in photomultiplier tubes. They can be used in transmission mode: the photoelectrons are emitted on the opposite side of the cathode layer from the photons (semi-transparent) or in reflection mode: the photoelectrons are emitted on the same side (opaque).

- **Cs-I** Cs-I is not sensitive to solar radiation ("solar blind"), its sensitivity cuts off at wavelengths *larger* than 200 nm and is exclusively used for ultraviolet detection. It is associated with MgF_2 crystals or synthetic silica to take advantage of their UV transmittance. To measure wavelengths smaller than 115 nanometers (cutoff of MgF_2), the window is removed and the Cs-I cathode is used in transmission mode.

- **Cs-Te** Cs-Te is also solar blind with a cutoff at 300 nanometers. Similarly to Cs-I, MgF_2 crystals or synthetic silica are used for the window material.

- **Sb-Cs** Sb-Cs is sensitive in the ultraviolet range and has the property of having a lower resistance than the bialkali photocathode. Therefore it is suited for applications where the light intensity to be measured is relatively high so a large current can flow in the cathode. It is also suited for applications where the photocathode has to be cooled (resistance becomes larger) and again avoid current limitation.

- **Bialkali (Sb-Rb-Cs, Sb-K-Cs)** So denoted because two types of alkali metals are used. Bialkali photocathode have a spectral range similar to the Sb-Cs photocathode but a higher sensitivity (QE) and a lower dark current.

- **High temperature, low noise bialkali (Sb-Na-K)** As with bialkali photocathodes, two kinds of alkali metals are used. The spectral response is identical that of the bialkali, but with lower sensitivity. This photocathode can withstand operating temperature up to 175°C, while photocathodes are normally not guaranteed for temperatures higher than 50°C. In addition, when used at room temperature, this photocathode exhibits very low dark current, which makes it useful in low-light measurements, such as photon counting where low noise is required.

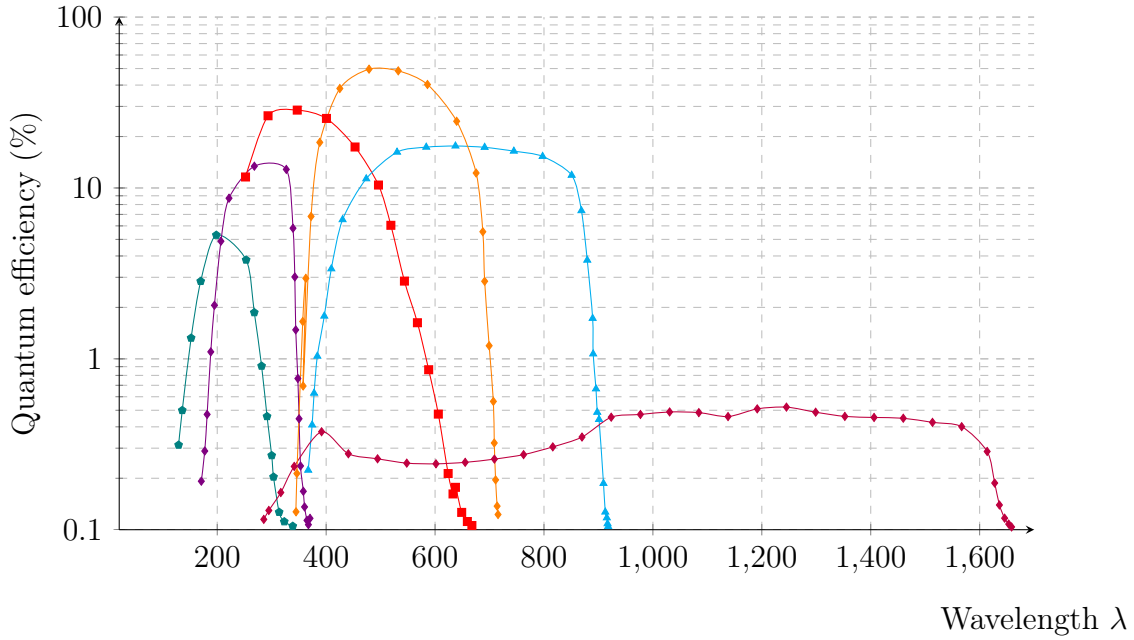


Figure 5.10 – Quantum efficiency versus wavelength of various photocathodes: Cs-Te (—●—), GaN (—◆—), SuperBialkali (—■—), GaAsP (—▲—), GaAs (—◆—), InP (—◆—). Quantum efficiency is the ratio of photoelectrons emitted from the photocathode to the number of incident photons, expressed in percent.

- **Multialkali (Sb-Na-K-Cs)** This photocathode uses three or more kinds of alkali metals. Due to its high sensitivity over a wide spectral range, from the ultraviolet to the near infrared region around 850 nanometers, this photocathode is widely used in broadband applications such as spectro-photometers.
- **Ag-O-Cs** Ag-O- Cs is sensitive from 400 to 1200 nanometers, it has a lower sensitivity in the visible region and is mostly used for near infrared detection.
- **GaAsP (Cs)** GaAsP crystals activated with cesium are used as transmission photocathode. This photocathode has a very high sensitivity in the visible region. However, if exposed to high light intensity, it is more likely to suffer degradation compared to other alkali photocathodes.
- **GaAs (Cs)** GaAs crystals activated with Cesium are also used as transmission photocathodes. They demonstrate a nearly flat, high sensitivity spectral

response from 350 to 850 nanometers. Similarly to GaAsP, this photocathode is sensitive to degradation under high light intensity exposure.

- **InP/InGaAsP (Cs), InGaAs (CS)** Field-assisted photocathode utilizing the PN junction formed by growing InP/InGaAsP or InP/InGaAs on a InP substrate. Applying a bias voltage to this photocathode allows reaching higher sensitivity at wavelengths up to 1.7μm, impossible with conventional photocathodes. This photocathode must be cooled down to -60°C to -80°C during operation to reduce the dark current.

Figure 5.10 shows the typical quantum efficiency versus the wavelength for some of the photocathode materials described.

5.5.2 Conclusion

Photocathode development is still an active field of research within the LAPPD project [12], [13], [14]. No clear choice or options have been taken since the photocathode is, in the end, most likely to be application dependent. For instance experiments measuring Cerenkov light in big neutrino detectors being more likely to use a bialkali photocathode and a time-of-flight experiment with ultra-short infrared laser pulses will be more interested in InP photocathodes.

5.6 Micro-channel plates

Micro-channel plates consist of an array of glass capillaries bundled together. Typically the diameter of the pores ranges from 6 to 40 microns. A 20 by 20 centimeter micro-channel plate with 40-micron pores is made of approximately ten million pores. The inner walls of the plate are processed to have the proper electrical resistance and coated with secondary emissive material. A photograph of an uncoated 8-inch square micro-channel plate is shown in Figure 5.11.

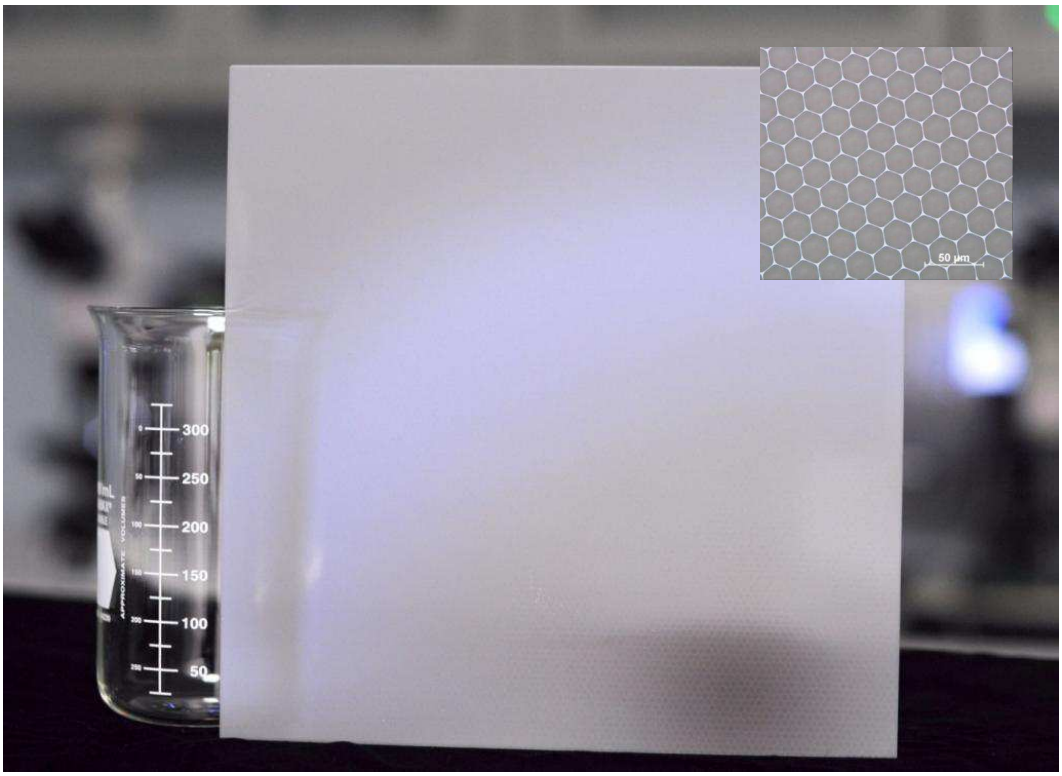


Figure 5.11 – Photograph of an uncoated 8 inch square micro-channel plate and detail of the pore structure, manufactured by Incom [15].

As described in the Introduction of this Chapter, the signal coming from the photocathode (a single to a couple of photoelectrons), is usually too small to be detected using conventional electronic readout systems. Therefore micro-channel plates are added in the path of the photoelectron in order to convert the single electron input into 10^5 - 10^6 electrons.

This must be done without compromising the time information of the signal, **the precise time when the photon hits the photocathode**.

Micropores are a perfect way to quickly amplify the photoelectron signal as shown in Figure 5.1. The other advantage of micropores is that the time information of the photoelectron is preserved during the amplification process, it will be discussed in the next Section.

5.6.1 Electron trajectography

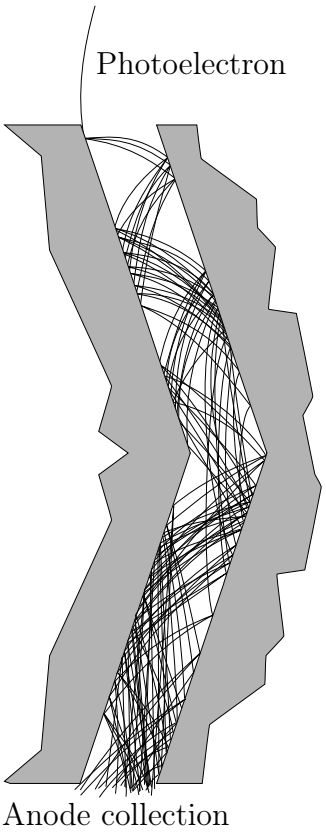


Figure 5.12 – Secondary emission amplification of electron in a micro-channel plate pore.

In order to collect the photoelectrons and secondary electrons efficiently on the anodes and minimize the transit time spread, the electron path in the channel must be optimized. The electron movement in the photomultiplier is set by the electrical field geometry inside the pore and by the applied voltage. As can be seen in Figure 5.12, simulating the electron trajectories inside a pore is a complicated problem. But the following empirical observations can be made about the different design and operating parameters of the micro-channel plate:

- **Pore size** The diameter of the pore size directly impacts the transit time of the electrons inside the pores. Smaller pores yield to smaller spread and better time performances.
- **Pore length** The length of the pore sets the number of times the photoelectron will hit the wall and create secondaries. A higher length gives more gain but also brings more dispersion and spread in the electron travel times. A pore too long will be more likely to saturate and take more time to recover from the amplification process.
- **Bias angle** The bias angle is necessary in order for the photo-electron to hit the wall of the pore. A bias angle too acute will decrease the uniformity of the

electron cloud inside the pore during the amplification process, while an angle too wide would simply let the photoelectron go through without amplification.

- **Voltage** The number of secondary electrons emitted increases with the high voltage applied across the pore. Therefore the gain of the micro-channel plate will increase with the applied voltage. However at high voltage the gain will be more dependent to variation in the high-voltage power supply, such as drift, ripples, temperature instabilities and load regulation.

5.6.2 Electron multiplication and gain

5.6.2.1 Secondary emission

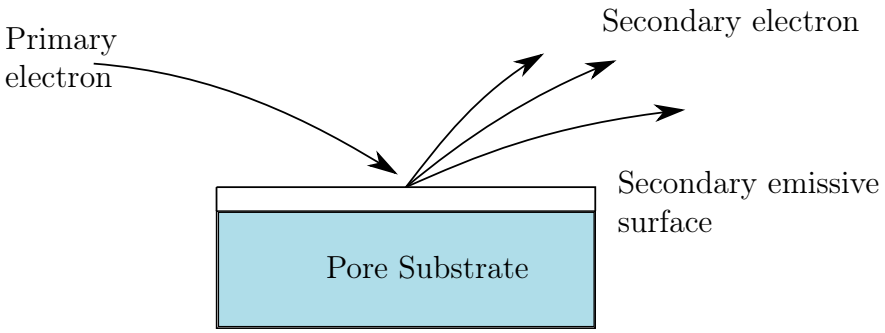


Figure 5.13 – Second emission amplification of electrons in a micro-channel plate pore.

As described previously, given enough energy, accelerated electrons hitting the micro-channel plate pores will trigger the emission of secondary electrons, effectively amplifying the signal. The inner wall of the micro-channels is coated with secondary emissive material in order to do so. The best and most commonly used materials are alkali antimonide, beryllium oxide (BeO), magnesium oxide (MgO), gallium phosphide (GaP) and Gallium phosphide (GaAsP). Figure 5.13 shows the principle of secondary emission when a primary electron hits the wall of the pore.

When a primary electron with initial energy E_p strikes the pore surface, δ secondary electrons are emitted, where δ is called the secondary emission ratio. Figure 5.14 shows the secondary emission ratio δ for various coating inside the pore, as a function of the accelerating voltage of the primary electrons. A higher accelerating voltage in the pore will create more secondaries (δ higher) and give more gain.

5.6.3 Gain

The current amplification or gain of a discrete secondary emission process is directly given by the parameter δ. If there are, on average, n stages of secondary emission for a single photoelectron, the total gain would be δ^n.

In a micro-channel plate the process of secondary emission happens continuously through the pore (see Figure 5.12), and the number of strikes on the walls of the pore is a continuous function of the length-to-diameter ratio $\alpha = L/d$ of a channel. Therefore the gain μ can be approximated as:

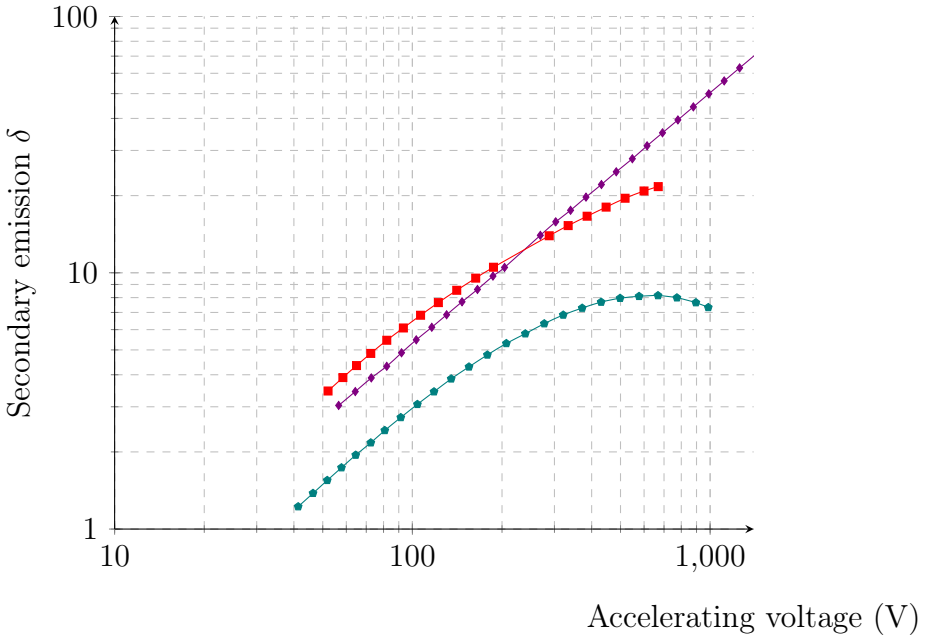


Figure 5.14 – Secondary emission ratio for Cu-BeO-Cs (—●—), K-Cs-Sb (—♦—) and GaP-Cs (—■—).

$$\mu \approx e^{\delta \times \alpha} \tag{5.4}$$

The exponential dependence of the gain is explained by the recursive nature of the amplification across the pores:

$$\begin{aligned} 1e^- &\rightarrow \delta e^- \\ \delta e^- &\rightarrow \delta^2 e^- \\ \delta^2 e^- &\rightarrow \dots \end{aligned} \tag{5.5}$$

The exponential gain dependence with the open area ratio α is explained by the fact that due to a higher number of secondary emissive steps in pores with longer length L, the gain will be higher. In pores with a smaller diameter d, the electrons will have a shorter mean free path between two hits on the sidewalls, increasing the gain as well.

In general a higher gain can be obtained using a higher α, however at higher α the micro-channel plate will turn on (be in the exponential gain range) for higher supply voltage.

5.6.4 Ion feedback

When the gain becomes higher than 10^4, the micro-channel plate becomes increasingly sensitive to noise effects on the supply voltage and stability becomes a serious problem.

At high energy, the created electron shower starts to ionize loosely bounded atoms, which are then accelerated towards the top of the channel. They can build up enough

energy to release electrons when they hit the channel wall. Then a second avalanche will be initiated. The "false" pulse will not only disturb the measurement, they could lead to permanent discharge and destroy the micro-channel plate. This effect is illustrated in Figure 5.15.

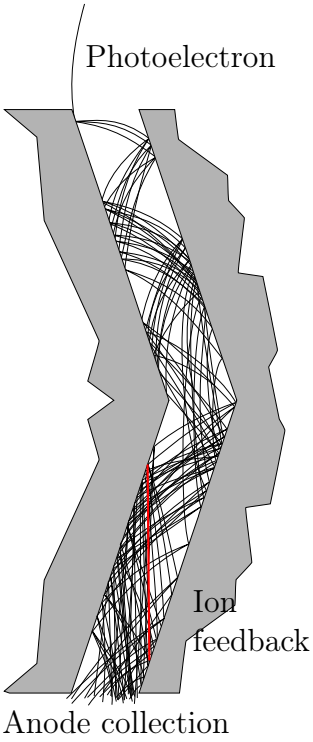


Figure 5.15 – Illustration of the ion feedback effect inside the micro-channel pore in the case of very high gain. The ion trajectory is represented in red. The electron cloud has ionized a loose atom inside the pore, which migrates to the upper part. Due to its high mass and positive charge, the ion has a straight, rising trajectory. The chevron geometry prevents the ion to escape to photocathode and damage it. Ions are more likely to be created in the lower section of the pores where the number of electrons is exponentially higher.

In order to avoid this effect, the ratio α is usually kept under 40 and higher gain is achieved stacking micro-channel plates together. During stacking the pores are oriented such as creating "chevrons"(see Figure 5.12) to avoid ions going from one plate to the other and to maintain a good uniformity of the electron cloud inside the pore.

5.6.5 Conclusion

The micro-channel plates used in the LAPPD project are much larger than typical commercial ones, causing difficulties to coat them uniformly with secondary emissive material. Using the technique of Atomic Layer Deposition - effectively coating the pore with one layer of atoms at a time - a very high uniformity and gain are obtained throughout the 8-inch square of the micro-channel plate. In order to have gain in the order of 10^6 electrons two stacked micro-channel plates are used. Their bias angle is

8° and the L/d ratio equal to 60.

The micro-channel plates are functionnalized for secondary emission by Atomic Layer Deposition at the Argonne National Laboratory [16].

5.7 Time characteristics

5.7.1 Introduction

As discussed in the previous Sections, the use of micro-channel plates gives very high gain while minimizing the transit time spread of the electrons from the photocathode to the anodes. This is achieved by using micron-size pores, with accelerating voltage up to 3000 Volts.

Fast photon detectors are the key to measure precisely relativistic particle time-of-flight.

Rise time and fall time The conventional definition of rise time is the time for the pulse to go from 10% to 90 % of the pulse amplitude. Due to the nature of the electron avalanche, the leading edge of a micro-channel plate detector response is negative. However, the leading edge is still considered as being the rise time.

Transit time spread The transit time spread characterizes the variation of time between the production of the photoelectron and the signal creation at the anode plane.

There are three possible sources of rise time spread in a micro-channel plate photo-detector:

- The cathode gap.
- The micro-channel plate stack.
- The anode gap.

5.7.2 Cathode gap

The cathode gap does not affect the rise time because electrons emitted from the cathode carry only a small amount of their initial energy, and hence a low initial velocity. Typically the wavelengths of the incident light are around 500 - 1000nm so the photons have an initial energy E_λ of about:

$$E_\lambda = \frac{hc}{\lambda} \sim 1 \ - \ 2eV \tag{5.6}$$

The work function W_e of the photocathode (energy required for the electron to go from a bounded state in the photocathode to a free state in the vacuum) takes a portion of this energy away, leaving the photoelectron with negligible energy $E = E_\lambda - W_e$ compared to the voltage across the cathode gap (typically 200V) $E \ll 200eV$.

Therefore, any variation due to non-uniformity on the photoelectron initial energy E will not contribute to a transit time spread in the cathode gap.

Due to the vertical field geometry in the cathode gap, any radial speed effect will not be corrected during the acceleration. This effect is illustrated in Figure 5.16, where the first photoelectron has a smaller radial speed $\overrightarrow{v_{y1}}$ than the second one $\overrightarrow{v_{y2}}$, therefore taking more time to cross the photocathode-to-micro-channel plate gap.

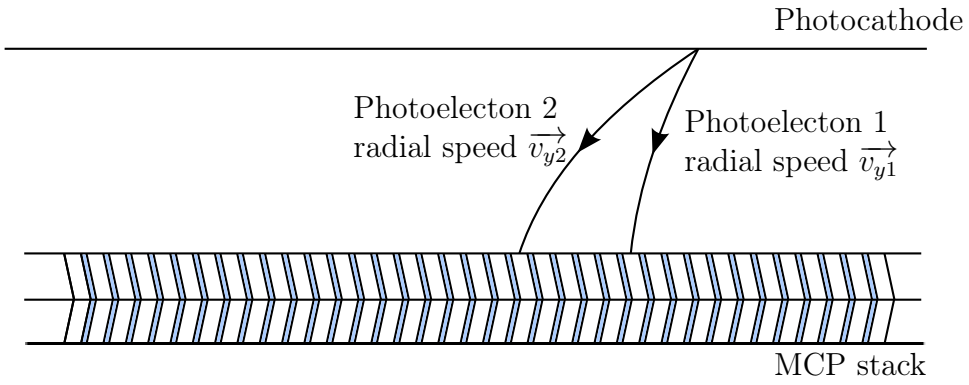


Figure 5.16 – Illustration of the effect of the radial speed distribution of the photoelectrons on their time of arrival to the micro-channel plates.

In order to minimize this effect, the gap distance between the photocathode must be made as small as possible and the cathode-to-micro-channel plate voltage high enough to minimize the transit time of the photoelectron to the plates and hence their radial shift.

Overall, due to the very high field in the cathode gap and its uniformity, the photocathode does not contribute in itself to the transit time spread of the photodetector. This shows, however, that the high voltage across the cathode gap plays an important role in ensuring low transit time spread and must remain very clean (no noise or ringing), uniform over the 8 by 8 inches of the photocathode surface and stable over time.

5.7.3 Micro-channel plate stack

The micro-channel plate is considered to be the main source of rise time spreading due to the different path lengths of the electrons through the pores. Using narrower pores is expected to reduce the path length variations. Longer pores would also allow the electrons for wider variations of path length and is therefore expected to increase the detector transit time spread.

Overall, a smaller open area ratio is expected to increase the timing performances of the micro-channel plate detector. We are going to study the effect of the pore geometry on the timing performances of the micro-channel plate detector in the following Sections.

5.7.4 First strike

Due to the bias angle of the micro-channel, the first impact location of the photo-electron in the tube depends on the region in the x-y plane where it entered the pore.

This effect is illustrated in Figure 5.17, where we consider two photoelectrons having their first interaction in the micro-channel plate pores at different depths. Photoelectrons hitting the pore sooner will have longer transit time in the pore and a higher number of interactions creating more secondary electrons - more gain.

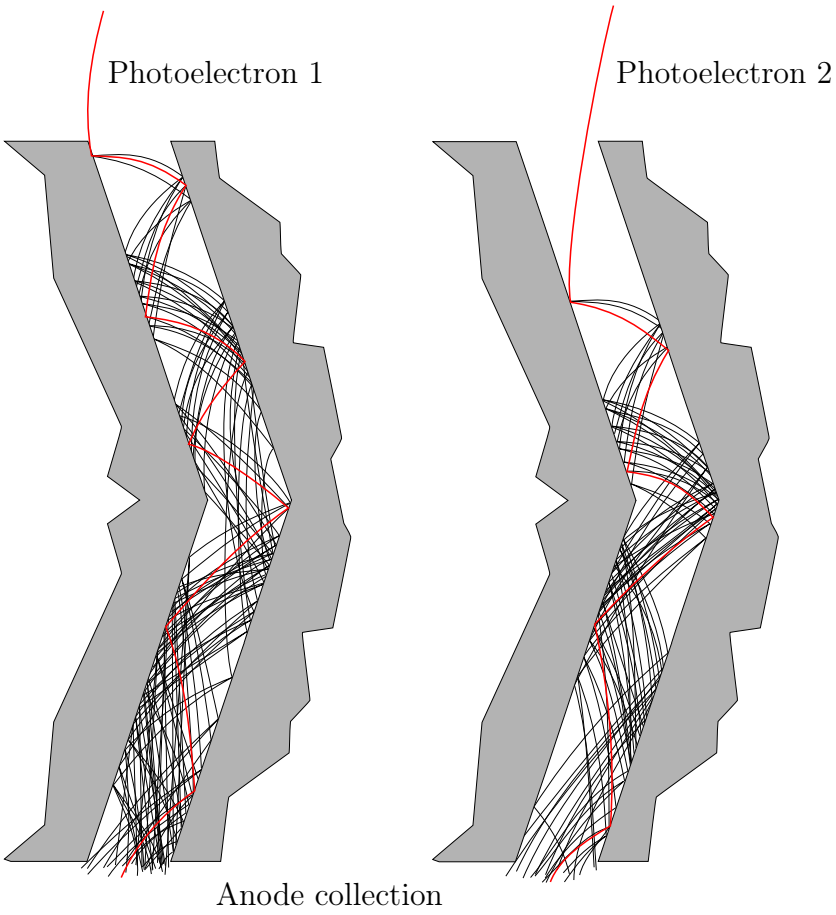


Figure 5.17 – Illustration of the first strike effect. Two photoelectrons hit the micro-channel plate pore at different depth and have different transit time spread. Photoelectron 1 will have a longer transit time, hitting the pore sooner, due to an increase number of bounces on the pore walls.

The percentage of variation in the transit time and gain of a given micro-channel plate can be very roughly estimated looking at the variation of the depth of interaction δ to the total length of the pore L. With d the pore size of the micro-channels and θ the bias angle of the pores we have:

$$\frac{\delta}{L} = \frac{d}{L \tan \theta} \tag{5.7}$$

For a micro-channel plate with a $\frac{L}{d}$ ratio of 60 and a bias angle of 8° we have a 11% variation in interaction depth in the pore.

5.7.5 Anode gap

The anode gap works similarly to the cathode gap, except that now the initial speed of the electrons in the gap is no longer negligible. If V is the voltage across the anode gap, the acceleration a of the electrons can be written as:

$$a = \frac{Ve}{dm_e} \tag{5.8}$$

With m_e the electron mass, e the electron charge and d the size of the gap. We can derive the transit time t_f of an electron with initial velocity v_0 in the anode gap of height d as:

$$t_f = \frac{dm_e}{eV}\left[\sqrt{v_0^2 + 2\frac{eV}{m_e}} - v_0\right] \tag{5.9}$$

This equation illustrates the dependence of the transit time to the initial velocity v_0 of the electron. A cloud of secondary electron with highly dispersed velocities exiting the pore will produce badly formed pulses.

However if we differentiate this equation with respect to the initial velocity v_0 we have:

$$\frac{dt}{dv_0} = \frac{dm_e}{eV}\left[\frac{v_0}{\sqrt{v_0^2 + 2\frac{eV}{m_e}}} - 1\right] \tag{5.10}$$

Which represents the dependence of the rise time upon the variation of the initial velocity. We can observe that increasing the high voltage V across the gap will reduce the effect of the velocity dispersion.

Because of the velocity spread and the arrival time dispersion of the secondary electrons on the anode, the anode gap is most responsible for the rise time of the micro-channel photo-detector.

5.7.6 Conclusion

In this Section we have presented the primary contributing effects due to geometry and parameters of the micro-channel plate detector, on its timing performances. We see that, in order to get the best timing performance, the mean free path of the electrons in the detector must be kept as small as possible, requiring small open area ratio for the pores and a small total height for the detector stack (anode and cathode gap). In the next Section, simulations of the effect of the variation of these parameters on the time resolution will be presented.

For the LAPPD project the choice of an open area ratio $\alpha = 60$ to 80 has been made, with 20 or 40 μm pores. This is a compromise between small pore size and

rigidity requirement for the 8 by 8 inch micro-channel plate. The total height of the detector stack is set at 1.2 cm.

5.8 Rise time

This Section presents a study of the effect of variations in the detector dimensions and polarisation voltages on the signal shape and on the expected time resolution, for a micro-channel plate detector.

5.8.1 Signal creation

Signal creation on the anode is due to the current induced by the moving electron charges in the gap from the back of the micro-channel plates to the anodes [17].

The electrons moving from the photocathode to the micro-channels and inside the micro-channels are actually "screened" by the metal layer used for high voltage polarization on the back side of the bottom micro-channel plate and do not contribute to signal formation.

5.8.2 Moving point charge

If we consider a moving point charge of charge q, accelerated from the micro-channel plate to the anodes with a speed v, the Shockley-Ramo theorem [17] states that the current I on the anode can be written as:

$$I = E(r) \times q \times v \tag{5.11}$$

Where $E(r)$ is the electric field at point r with **unit** potential applied on the anode, with all other parts of the detector grounded. This formula is only valid in the limit of the electrostatic approximation, where the retardation effects are negligible throughout the volume in which the charges move.

In our case, we can approximate the anode array and bottom side of the micro-channel plate as an infinite parallel plates separated by a gap distance d, with the electrical field $E(r)$ at point r with a unit potential applied expressed as:

$$E(r) = \frac{-1}{d} \tag{5.12}$$

The validity of this equation can be verified doing the integration of the total current:

$$\int_{t_0}^{t_f} I dt = -\frac{q}{d} \int_{t_0}^{t_f} v dt = q \tag{5.13}$$

The total charge deposited on the anode corresponds to the amount of charge q that was accelerated in the gap.

For an electron accelerated in a gap d with a voltage difference $V$and initial speed v_0 at t_0, Equation 5.8 gives the speed v of the electron:

$$v(t) = \frac{qV}{m_e d}(t - t_0) + v_0 \qquad t \in \left[0, \frac{dm_e}{Vq}\left[\sqrt{v_0^2 + 2\frac{Vq}{m_e}} - v_0\right]\right] \tag{5.14}$$

We can also derive the maximum speed v_{max} of the electrons for $t = t_f$:

$$v_{max} = \sqrt{v_0^2 + 2\frac{Vq}{m_e}} \tag{5.15}$$

In that case the generated current on the anode can be written as:

$$I = -\frac{q^2 V}{m_e d^2}(t - t_0) - \frac{q v_0}{d} \qquad t \in \left[0, \frac{dm_e}{Vq}\left[\sqrt{v_0^2 + 2\frac{Vq}{m_e}} - v_0\right]\right] \tag{5.16}$$

And the maximum current on the anode I_{max} is:

$$I_{max} = \frac{q}{d}\sqrt{v_0^2 + 2\frac{Vq}{m_e}} \tag{5.17}$$

We can make several observations from this Equation:

- A smaller anode gap achieves higher levels of current excitation on the anode.
- A higher voltage V across the gap achieves higher levels of current excitation on the anode.
- The formula in 5.17, being similar to Equation 5.9, differentiating with respect to v_0 leads to the same conclusion: a higher voltage V across the gap will reduce the gain variation on I_{max} due to velocity dispersion.

This crude model leads to an important observation: reducing the gap dimension d or increasing the gap voltage V will make the pulse shorter with a higher maximum current I_{max} but will not increase the total energy - or area under the pulse shape - of the signal. The typical pulse shape for a perfect anode is shown in Figure 5.18. The current excitation due to the moving charge in the gap is directly proportional to the charge q. A simple way to increase the current is to increase the number of charge, which will be considered in the next Section.

5.8.3 Moving electron cloud

In micro-channel plate detectors, between 10^5 to 10^7 electrons are typically created, which are simultaneously generating current on the anode.

If one considers all electrons as independent, we can apply the principle of superposition and if v_i is the speed of the n electrons we have:

$$I = \sum_{i=1}^{n} -\frac{v_i(t)e}{d} \tag{5.18}$$

If we have t_{i0} the time of emission of the electron i in the anode gap and we have a cloud of n electrons, we can define their mean time of emission t_{mean} as:

$$t_{mean} = \frac{1}{n}\sum_{i=1}^{n} t_{i0} \tag{5.19}$$

Similarly we can define their mean intial speed v_{mean} by:

$$v_{mean} = \frac{1}{n}\sum_{i=1}^{n} v_{i0} \tag{5.20}$$

With v_{i0} the initial speed of electron i. Using the definition of t_{mean} and v_{mean} and for a large number we can write the current I of a cloud of n electrons as:

$$I = -\frac{ne^2V}{m_e d^2}(t - t_{mean}) - \frac{nev_{mean}}{d} \tag{5.21}$$

With this equation we can identify three primary fluctuation effects on rise-time, gain and transit time, in the detector response.

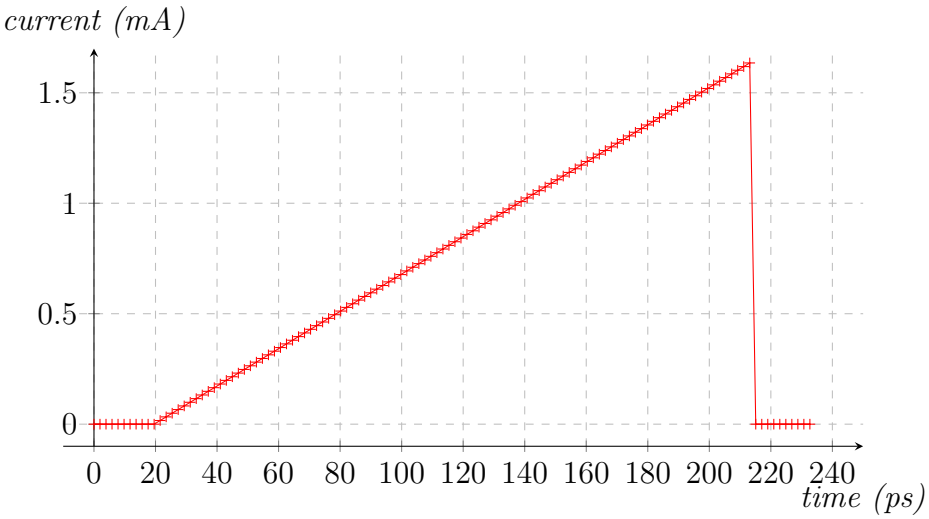


Figure 5.18 – Current on the anode (—+—) done for $n = 10^6$ electrons in a 1mm gap biased with $V = 200V$. The simulated pulse duration is 200ps.

5.8.3.1 Rise-time fluctuation

Equation 5.21 shows that the current response is a linear function of slope s:

$$s = -\frac{ne^2V}{m_e d^2} \tag{5.22}$$

This slope can be directly associated with the rise-time of the micro-channel plate response. We can observe that three parameters can improve the rise time of pulse:

- Higher accelerating voltage V.

- Higher number of secondary electrons n.
- Smaller pore size d

If we have a variance σ_n^2 on the number of electrons in the cloud (due to first strike or statistical fluctuation in the pores for example) we have a variance σ_g^2 on the rise time of the pulse written as:

$$\sigma_g^2 = \sigma_n^2 \left(\frac{e^2 V}{m_e d^2} \right)^2 \tag{5.23}$$

In order to get a lower dispertion (lower σ_g) for the rise-time, a higher number of electrons in the cloud will help statistically. Additionally we see that fluctuations or instabilities on the high voltage V across the gap will also create rise-time dispersions.

5.8.3.2 Transit-time fluctuations

The y-intercept anode current equation 5.21, can be assimilated to the start time t_{start} of the pulse on the anode, and:

$$t_{start} = t_{mean} \tag{5.24}$$

t_{mean} represents the electrons mean time-of-arrival in the micro-channel plate pores. Therefore, since the current signal on the anode starts to form as soon as the electron start moving out of the micro-channel pore, the transit time does not depend on the electrons transit time inside the anode gap. The detector transit time spread is only dependent on the transit time inside the micro-channels.

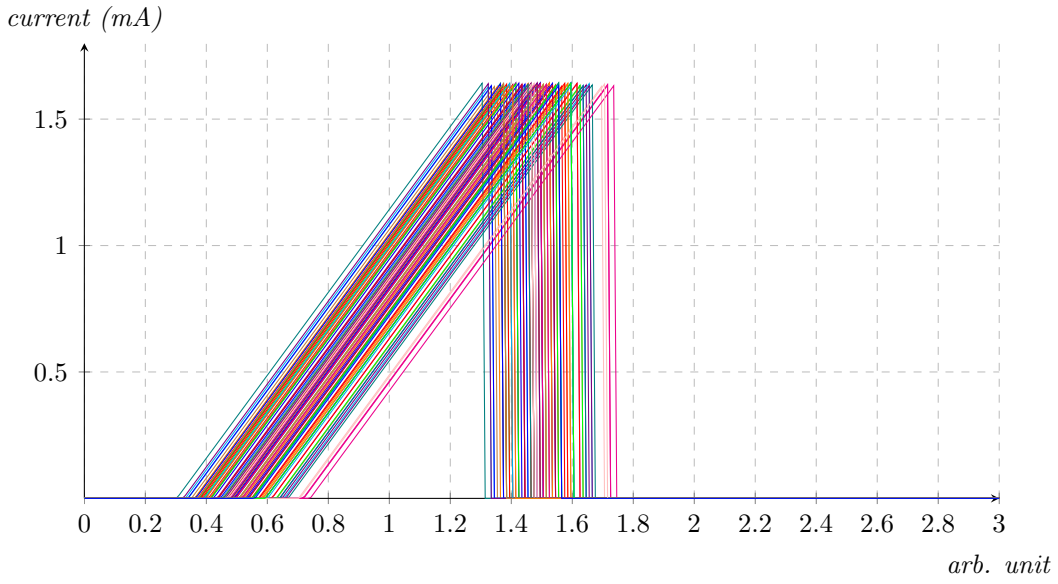


Figure 5.19 – Simulation of the effect of the transit time spread noise on the anode current. In the simulation we have set the noise level at 10% of the electrons transit time in the gap.

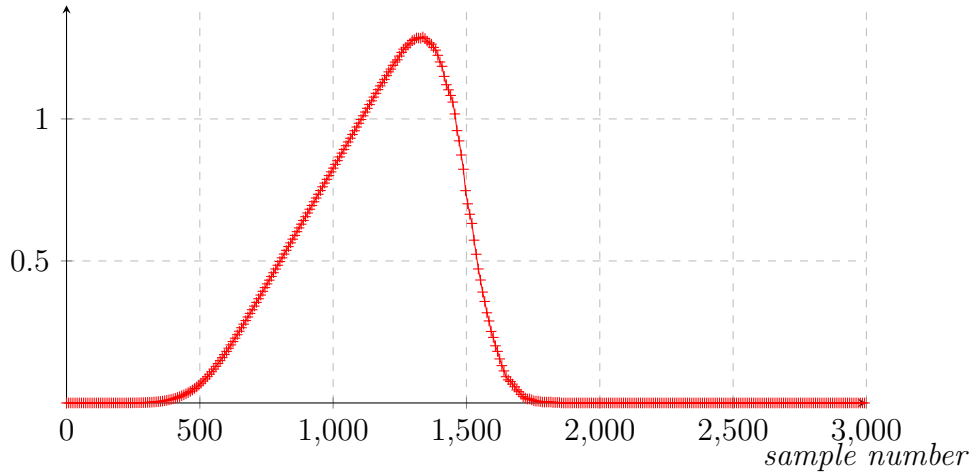


Figure 5.20 – Current on the anode (—+—)due to the transit time spread in the micro-channel plates.

5.8.3.3 Gain

The formulae giving the gain of the pulse G in case of a high number of electrons can be approximated summing the maximum current I_{max} produced by the n secondary electrons:

$$G \approx \sum^{n} \frac{e}{d} \sqrt{v_{0i}^2 + 2\frac{Ve}{m_e}} \tag{5.25}$$

Here we find again the result we stated previously using Equation 5.10: the gain fluctuation due to the initial speeds v_{0i} of the electrons in the gap can be reduced using a higher voltage V, though only in the square root.

5.8.4 Fall time

We can see in Figure 5.18 that the fall time of the current on the anode is non-physical, dropping to zero when the electron reaches the metal, apart from smearing due to noise effect (see Figure 5.20), the fall time is not controlled by the physics in the gap.

The fall time and the shape of the fall time are principally defined by the capacitive effects on the anodes and the impedance of the transmission line carrying the signal out of the detector (see Section 5.9).

Indeed the very fast electrical excitation of the electrons on the anodes have a very wide-band spectrum. Any attenuation due to the anode in the spectrum of the signal will severely distort (phase shift) and attenuate it.

5.9 Anodes

The anode is the structure used to collect and transmit signals from the micro-channel plate to outside the detector. For the LAPPD photo-detector project a stripline array geometry was used. It consists of 30 striplines, uniformly spread on an 8 inch wide Borofloat 33 glass substrate. A picture of the anode geometry is showed in Figure 5.21

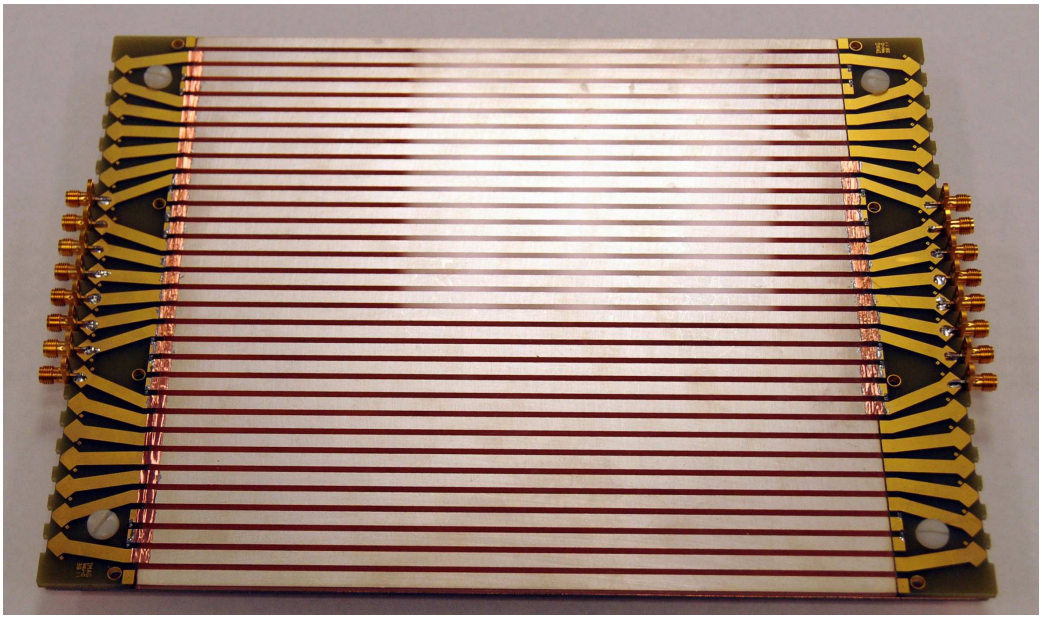


Figure 5.21 – Photograph of the 30-strip anodes array used to collect and transmit current pulses from the 8 inch LAPPD photo-detector. The anode array has been connected to two side cards (on the picture) used for characterization of its radio-frequency properties.

5.9.1 Introduction

As we saw in Section 5.8, the charge cloud emerging from the micro-channel plate pore holds the time information generated by the initial photoelectron. Additionally, because of the micro-channel plate pore segmentation, the electron cloud contains also spatial informationregarding where the photoelectron hit the photocathode.

The role of the anode array is to transmit the time and space informations of the generated current pulse outside the detector. On each stripline, pulses created by the charge excitation propagate in opposite directions towards the end of the lines, where they are digitized using waveform sampling. The coupling to the waveform sampling electronics is shown in Figure 5.22

5.9.2 Measuring position and time

From the digitized pulse waveforms sampled at each end of the stripline, we can determine the time, position, total charge and pulse shape generated by the photoelectrons. The longitudinal spatial location is determined from the difference in

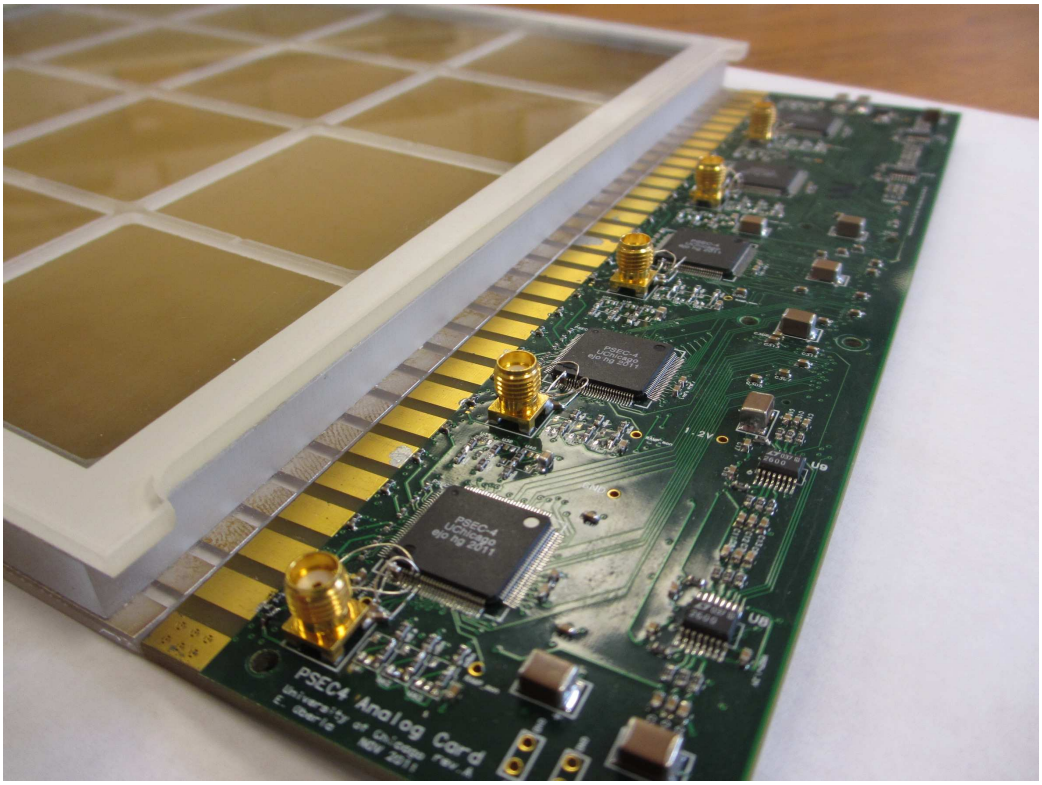


Figure 5.22 – Photograph of one end of the 30 strip anodes array connected to the waveform sampling readout electronics (5 PSEC4 chip).

time measured at the two ends of the strip. One-dimensional stripline allows to considerably reduce the number of readout channel compare to anode pad arrays while preserving decent space resolution.

Another advantage of one-dimensional striplines is to make possible signal pick-up on the sides of the detector, removing the need for individual feedthrough on the 400 cm^2 back of detector (which would have made it extremely fragile and hard to manufacture).

5.9.2.1 Spatial resolution

The spatial location of the photoelectron excitation along the strip is determined by the difference in times measured at both ends of the strips. This is illustrated in Figure 5.23. If t_1 and t_2 are the measured times at both ends, v_p the velocity of the signal propagating along the striplines, the position of the hit x_{hit} on the stripline is simply given by:

$$x_{hit} = \frac{t_1 - t_2}{v_p} \tag{5.26}$$

In the transverse direction the resolution is determined by the strip spacing. So, if the effective detector size is s in the transverse direction and we have p anodes, the position in the transverse direction y_{hit} is given by:

$$y_{hit} = \frac{s}{p} \tag{5.27}$$

In reality, when the electron cloud lands at the edge of one stripline, there will be charge sharing effects to the neighboring strips and by measuring the charge of the waveforms at the end of every strips sub-stripline pitch resolution can be achieved.

5.9.2.2 Time resolution

Times measured at both ends of the striplines t_1 and t_2 differ from the time of the excitation t_0 as they include the propagating time in each direction of the stripline. If L is the length of the stripline:

$$t_1 + t_2 = \frac{x_{hit}}{v_p} + t_0 + \frac{L - x_{hit}}{v_p} + t_0 = 2t_0 + \frac{L}{v_p} \tag{5.28}$$

The mean of the times measured at both ends of the stripline $\frac{t_1+t_2}{2}$ gives the information on the photoelectron time-of-arrival with a fixed offset.

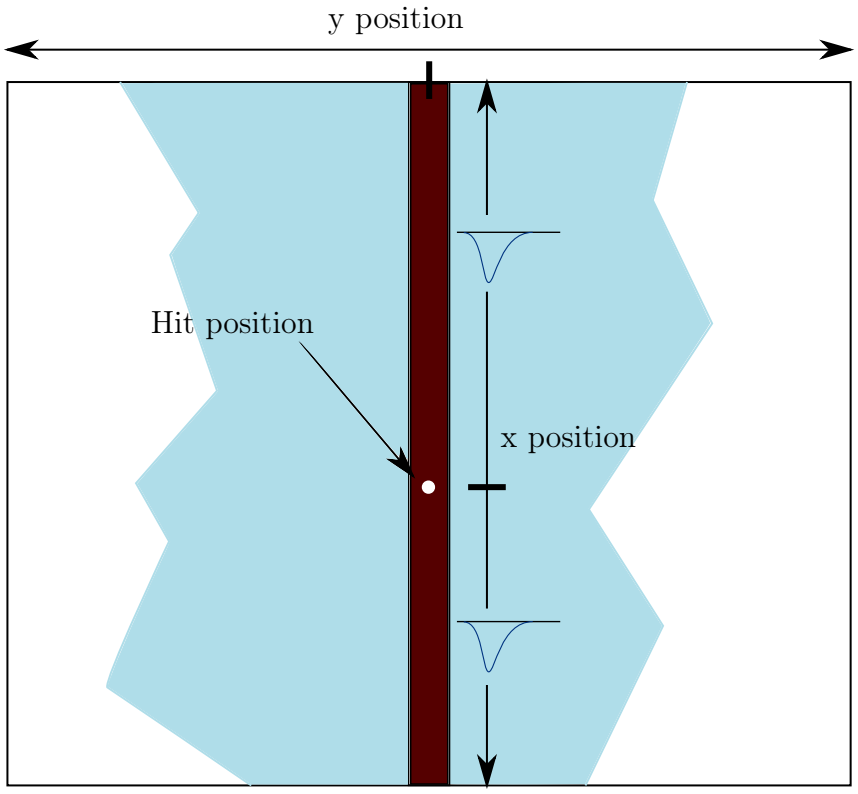


Figure 5.23 – Illustration of the position resolution using anode striplines. In the y-direction the position is given by the strip location. In the x-direction the position is given by the time of arrival of the signals at both ends of the stripline.

5.9.3 Micro-strip lines

The geometry of the anode is in fact an array of micro-strip lines. The basic schematic of a micro-strip line is shown in Figure 5.24. It consists of a conductive strip of width w and a wider ground plane, separated by a dielectric layer (also called substrate) of thickness h.

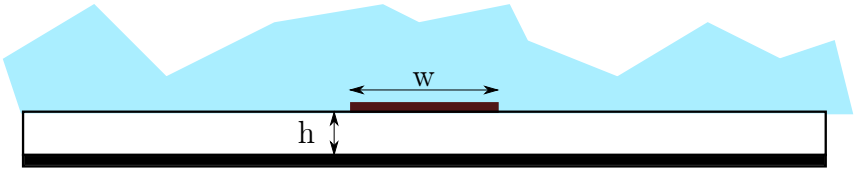


Figure 5.24 – Schematic of the typical geometry of a micro-strip line. The width of the line is w and the substrate height h.

Transmission lines with conductors embedded in an inhomogeneous dielectric medium cannot support a pure Transverse Electro-Magnetic (TEM) mode. This is the case for the micro-strip transmission line, where part of the field is in the air and part is between the conductor and a ground plane. However it can be shown that when the dielectric substrate is thin $h \ll \lambda$, the transverse field is dominant and called quasi-TEM [18].

5.9.3.1 Effective permittivity

The micro-strip effective permittivity ϵ_e is defined by [18]:

$$\epsilon_e = \frac{c}{v_p} \tag{5.29}$$

where c is the speed of light. Using Electromagnetic field solver the micro-strip effective permittivity can be fitted to the approximate formula [18]:

$$\epsilon_e = \frac{\epsilon_r + 1}{2} + \frac{\epsilon_r - 1}{2} \frac{1}{\sqrt{1 + 12^{h}/_{w}}} \tag{5.30}$$

with an error of less than 1% for $1 \leq {}^{w}/_{h} \leq 100$ and $1 \leq \epsilon_r \leq 128$.

5.9.3.2 High Frequency properties

Detailed analysis [19] of the properties of micro-strip lines yields to the following formula for the frequency-dependent permittivity of a micro-strip line:

$$\epsilon_e(f) = \epsilon_r - \frac{\epsilon_r - \epsilon_e(0)}{1 + (f/f_a)^m} \tag{5.31}$$

with the corner frequency f_a defined by:

$$f_a = \frac{f_b}{0.75 + (0.75 - 0.332\epsilon_r^{-1.73})^{w}/_{h}} \tag{5.32}$$

$$f_b = \frac{47.746 \times 10^6}{h} \times \mathcal{F}(w/h) \tag{5.33}$$

with $\mathcal{F}(w/h)$ a function of the ratio w/h. Here, the point is to draw attention to the corner frequency f_a being proportional to $\frac{1}{h}$.

$$f_a \propto \frac{1}{h} \tag{5.34}$$

In micro-channel plate photo-detectors, the thickness of the anodes h cannot be made arbitrarily small, as it would compromise the solidity of the detector. Therefore, given the anode material and impedance, this typically limits the highest frequency that can propagate along the striplines.

5.9.4 Anode RF connections

5.9.4.1 RF connections

In order to characterize the anode bandwidth, the frequency response has been measured using an Agilent HP8753E network analyzer [20]. The transition of the electromagnetic field between the coaxial geometry of the SMA cables to the micro-strip creates distortions and reflections of the signal. [4].

Figure 5.25 – Close-up photograph of the improved SMA connection to the micro-strip line.

In order to incrementally adapt the field from one geometry to the other, a manual tuning of the impedance of the transition region is done using adhesive copper tape. During the operation, by monitoring the signal transmission and reflection in the time domain on a network analyzer, one can identify the location of impedance mismatches and make appropriate additions (more capacitance) or subtractions (more inductance). This is shown in Figure 5.25, where we can see a close-up view of the connection from the SMA connector to the stripline.

5.9.4.2 Fanout board

In order to further characterize the anode array, in particular the crosstalk, and higher excitation modes of the coplanar striplines, a PCB fanout board has been fabricated to allow easier parallel connections. Due to the high striplines density, only 3 groups of 6 adjacent striplines were connected, the others were terminated on board. On the board, the SMA to micro-strip field transition was matched using the geometry extracted from previous work (Figure 5.25).

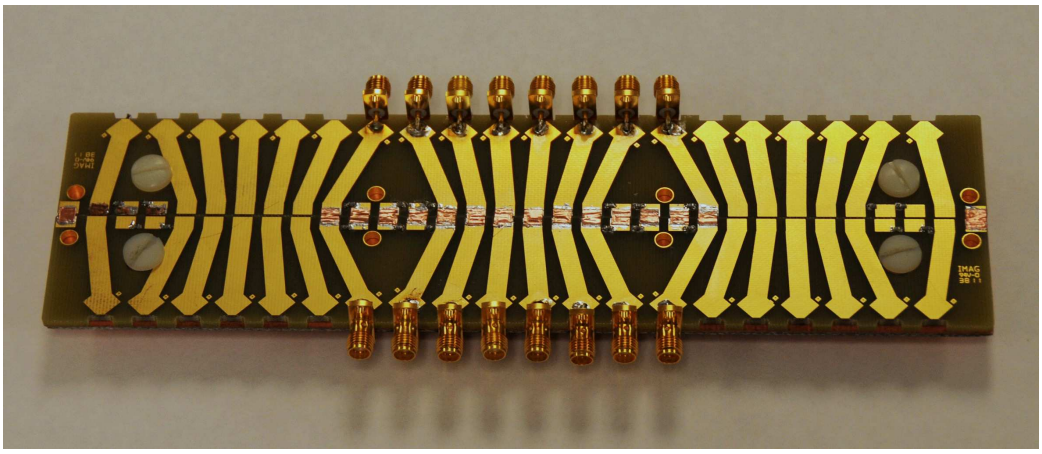


Figure 5.26 – Photograph of two fanout boards attached together (for calibration measurements). Near the SMA connectors, the metal geometry inspired from the previous copper work (Figure 5.25) is visible.

5.9.5 Anode RF measurements

The bandwidth of the anode geometry was measured using the fanout board connected by soldering small strips of copper from the micro-strip to the fanout board, as shown in Figure 5.21.

Two types of anodes have been tested: a "40-strip" anode and a "30-strip". The bandwidth measured in decibel is shown in Figure 5.27. We can see an improvement going from the "40-strip" geometry to the "30-strip", but the bandwidth still remains limited to 2 GHz. Several reasons can be held responsible for the limitation:

- **Reflections and distortions** Despite the careful work done for field and impedance matching, going from the cables to the anodes, significant attenuation was observed at the transitions. The data shown in Figure 5.27 does not account for these attenuations.
- **Substrate height** By far the most limiting factor here, the relatively thick substrate, lowers the frequency at which, higher order, non-TEM modes appear in the micro-strip array.

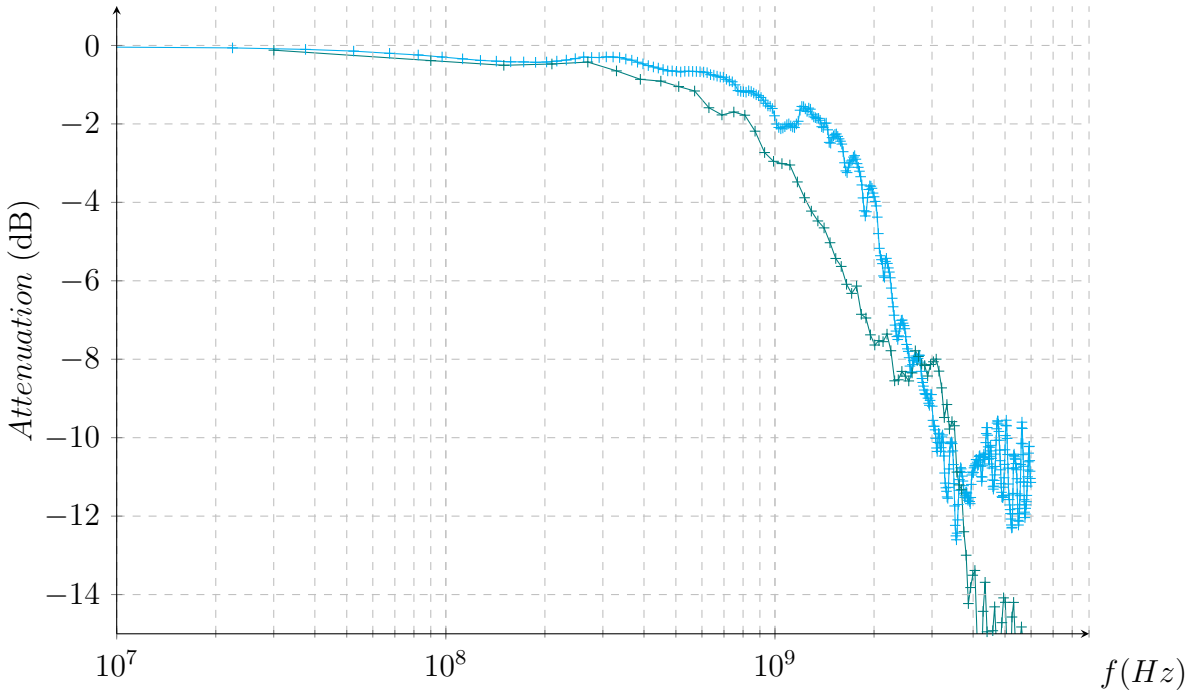


Figure 5.27 – Bandwidth of the anode geometry versus the frequency for 40-strips anodes (—+—) and 30-strips anodes (—+—).

5.9.6 Anode RF limitation

The frequency response of the anodes shown in Figure 5.27 can be seen as a simple low-pass filter, therefore, it will cutoff all the higher frequency components of the current excitation.

A simulation of the bandwidth limitation effect on the current excitation is shown in Figure 5.28. The simulation is done with a low-pass filter of order 1 and cutoff frequency 2 GHz.

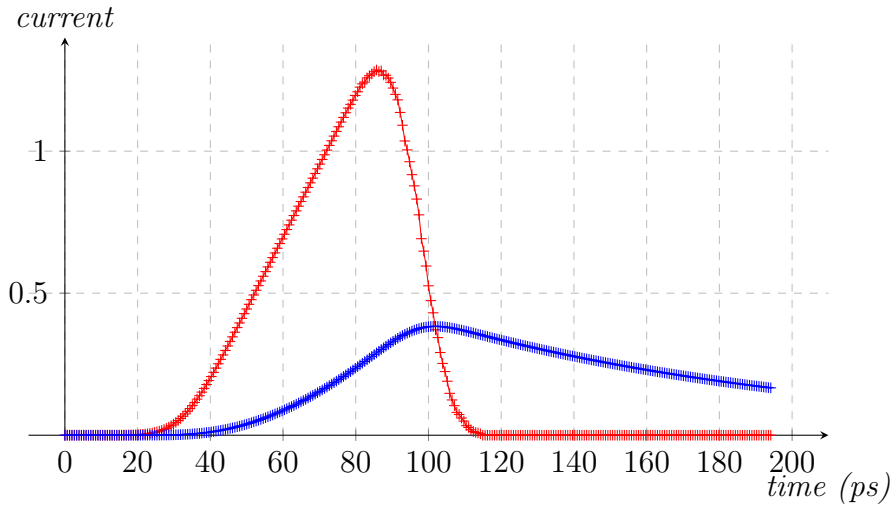


Figure 5.28 – Simulation of the pulse degradation due to the anode frequency limitation. The current excitation of the anode is plotted in red (—+—) and the response of the anode is plotted in blue (—+—) for a cutoff frequency of 2 GHz.

Interestingly, we note that the bandwidth limitation does not degrade too much the signal rise time, while the fall time is more seriously affected. A second simulation, done with a cutoff frequency set at 8 GHz, confirms this observation. In Figure 5.29, the fall time is more seriously affected by the bandwidth limitations.

This confirms the observation made Section 5.8: for micro-channel plate detectors, the fall time is limited by the anode properties. In conclusion, in order to preserve as much as possible the shape of the signal coming out of the micro-channel detector, it is essential to study carefully the anode RF properties so as to preserve the timing information.

As of today, the best reported bandwidth of the LAPPD photo-detector anodes is 2.6 GHz. Getting a higher bandwidth is still an on-going process at the University of Chicago and recent developments suggests that a bandwidth as high as 10 GHz could be reached, using a novel inside-out micro-strip geometry [21].

5.10 Conclusion

The LAPPD project, of which I was member for two years, in particular during the first year of this thesis work, is still an on-going project.

Most of the design choice were oriented as a tradeoff to develop cheap, robust, large-area photo-detectors, (for example by the use of cheap glass substrate: the Borofloat33 from Schott, or a large anode readout system) and trying to get the best timing (searching for the best pore geometry in the micro-channel plates or trying to get the fastest bandwidth out of the anode array).

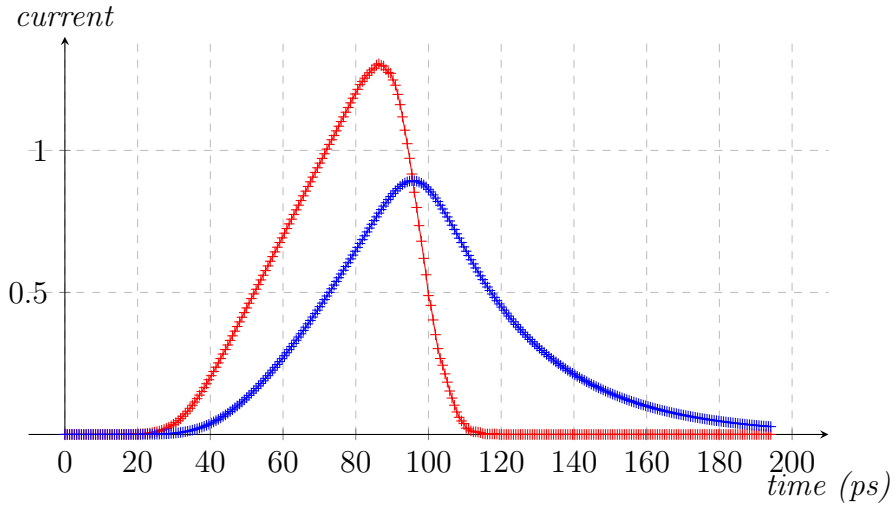


Figure 5.29 – Simulation of pulse shape degradation due to the anode frequency limitations. The current excitation of the anode is plotted in red (—+—) and the response of the anode is plotted in blue (—+—) for a cutoff frequency of 8GHz.

Today this project is very successful, succeeding in the first sealing of a 20 cm by 20 cm air-tight functional photo-detector. Thus, opening new perspectives to the instrumentation of large detector areas in particle physics. Demonstrated timing performance [2] and the excellent resolutions obtained [3] allows us to envisage it as the perfect candidate for a picosecond timing and time-of-flight photo-detector.

Bibliography

[1] KK Hamamatsu Photonics, Editorial Committee, et al. *Photomultiplier tubes-basics and applications*. *Edition 3a*, 2006.

[2] Camden Ertley, John Anderson, Karen Byrum, Gary Drake, Henry Frisch, J-F Genat, Harold Sanders, and Fukun Tang. *Development of Picosecond-Resolution Large-Area Time-of-Flight Systems*. *Nuclear Science, IEEE Transactions on*, 56(3):1042–1045, 2009.

[3] OHW Siegmund, JB McPhate, JV Vallerga, AS Tremsin, SR Jelinsky, and HJ Frisch. *Novel large format sealed tube microchannel plate detectors for Cherenkov timing and imaging*. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 639(1):165–168, 2011.

[4] Hervé Grabas, Razib Obaid, Eric Oberla, Henry Frisch, Jean-Francois Genat, Richard Northrop, Fukun Tang, David McGinnis, Bernhard Adams, and Matthew Wetstein. *RF strip-line anodes for Psec large-area MCP-based photodetectors*. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 2013.

[5] Bernhard Adams, Andrey Elagin, Henry Frisch, Razib Obaid, Eric Oberla, Alexander Vostrikov, Robert Wagner, and Matthew Wetstein. *Measurements of the gain, time resolution, and spatial resolution of a 20× 20cm MCP-based picosecond photo-detector*. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 2013.

[6] Eric Oberla, Jean-Francois Genat, Hervé Grabas, Henry Frisch, Kurtis Nishimura, and Gary Varner. *A 15 GSa/s, 1.5 GHz Bandwidth Waveform Digitizing ASIC*. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 2013.

[7] E Oberla, H Grabas, M Bogdan, H Frisch, JF Genat, K Nishimura, G Varner, and A Wong. *A 4-Channel Waveform Sampling ASIC in 0.13 μm CMOS for front-end Readout of Large-Area Micro-Channel Plate Detectors*. *Physics Procedia*, 37:1690–1698, 2012.

[8] Michael Cooney, Matt Andrew, Kurtis Nishimura, Larry Ruckman, Gary Varner, Hervé Grabas, Eric Oberla, Jean-Francois Genat, and Large Area Picosecond Photodetector Collaboration. *Multipurpose Test Structures and Process Characterization using 0.13 μm CMOS: The CHAMP ASIC*. *Physics Procedia*, 37:1699–1706, 2012.

[9] Henry Frisch, Jean-Francois Genat, Hervé Grabas, Chien-Min Kao, Chin-Tu Chen, Heejong Kim, and Fukun Tang. *Use of flat panel microchannel photomultipliers in sampling calorimeters with timing*, March 9 2011. US Patent App. 13/044,442.

[10] John Anderson, Klaus Attenkofer, Eric Delagnes, Henry Frisch, Jean-Francois Genat, Herve Grabas, Mary K Heintz, Edward May, Samuel Meehan, Eric Oberla, et al. *Considerations about large area –low coest fast imaging photodetectors*. 2010.

[11] *Schott*. http://www.schott.com/.

[12] OHW Siegmund, JB McPhate, AS Tremsin, SR Jelinsky, HJ Frisch, J Elam, A Mane, R Wagner, MJ Minot, J Renaud, et al. *20 cm Sealed Tube Photon Counting Detectors with Novel Microchannel Plates for Imaging and Timing Applications*. *Physics Procedia*, 37:803–810, 2012.

[13] Zikri Yusof, Klaus Attenkofer, Marcel Demarteau, Henry Frisch, Joe Gregar, Sharon Jelinsky, Seon Woo Lee, Jason McPhate, Richard Northrop, Alexander Paramonov, et al. *Development of an Alkali Transfer Photocathode for Large Area Microchannel Plate-Based Photodetectors*. *Physics Procedia*, 37:733–739, 2012.

[14] Ryan Dowdy, Klaus Attenkofer, Henry Frisch, Seon Woo Lee, Xiuling Li, and Steve R Ross. *Development of Ultra-Thin GaAs Photocathodes*. *Physics Procedia*, 37:976–984, 2012.

[15] *Incom*. http://incomusa.com/.

[16] Anil U Mane, Qing Peng, Matthew J Wetstein, Robert G Wagner, Henry J Frisch, Oswald HW Siegmund, Michael J Minot, Bernhard W Adams, Matthieu C Chollet, and Jeffrey W Elam. *A novel atomic layer deposition method to fabricate economical and robust large area microchannel plates*. In *SPIE Defense, Security, and Sensing*, pages 80312H–80312H. International Society for Optics and Photonics, 2011.

[17] William Shockley. *Currents to conductors induced by a moving point charge*. *Journal of Applied Physics*, 9:635, 1938.

[18] IJ Bahl and DK Trivedi. *A designer's guide to microstrip line*. *Microwaves*, 16(5):174–176, 1977.

[19] T.C. Edwards and B. Steer. *Foundations of interconnect and microstrip design*. John Wiley, 2000.

[20] *Agilent*. http://www.agilent.com.

[21] InnoSys. *High Energy Physics Detectors and Instrumentation; A High Bandwidth LAPPD Anode*. Technical Report DE-FOA-0000760, Departement of Energy - Small Business Technology Transfer, 2013.

6 Measuring picosecond time

Contents

6.1 Introduction

In this Chapter we study the mean of extracting the time information from a signal of interest. Firstly we are going to describe what the time of an event is and how to reference it. Then we will present two algorithms used to extract the time information from a signal and discuss their relative performance in terms of precision. In the end we will present the key parameters of arbitrary signals and their impact on the ultimate time resolution. For this case a MATLAB [1] simulation extracting the time information will be presented.

6.2 Absolute time versus relative time

There are two distinct ways of measuring the time of an event, absolute time and relative time.

6.2.1 Absolute time

Absolute time is often the first representation of time that comes to mind. It corresponds to the dating process, such as stating *it is Dec 3rd 2013 12.36PM*. The observer defines a given, precise moment. In reality the absolute time is always measured with respect to a known reference, dating for example uses *Anno Domini* in the Julian calendar.

6.2.2 Relative time

Relative time is the representation of elapsed time, corresponding to a duration. In physics experiments time is always measured relative to a known reference:

- Beginning of universe in cosmology (not a precise reference but sufficient for the scales considered).
- Earth formation in geology.
- Stopwatch in races.
- Digital oscillator or reference clocks in electronics.

In our case, time is measured relative to a known, stable clock provided by the physics experiment we are involved with, the signal of interest being sampled synchronously to that clock. In the rest of this Section we will assume that the reference clock used for sampling is perfect, providing an absolutely clean reference.

For our particular application, where we want to measure time values below 10 picoseconds, the time reference employed should at least be 1 order of magnitude better (*i.e.* $\leq 1ps$) which is commonly achievable using today's precise oscillators and clock distribution systems [2].

6.3 Measuring time of arrival

One of the challenges of this work is to define a method allowing precise (sub 10 picoseconds) time-of-arrival measurement of an event in the detector. This chapter reviews some of the techniques available to extract the time of sampled signals from a detector and their precision.

We assume that the signal is sampled with respect to a very precise synchronous clock. A suficient number of samples is taken to cover the full length of the signal of interest. The digitized signal is processed by a computer algorithm to extract its amplitude and time.

There are basically two methods:

- Reconstructing the signal and applying a time extraction algorithm on the reconstructed waveform. Many implementations are possible but they are all essentially based upon threshold crossing measurements, whether it is a single threshold crossing (threshold or constant fraction) or multiple thresholds (linear interpolation or splines) [3].
- Direct extraction of time using estimators such as cross-correlation [4], [5].

Both options will be investigated.

6.3.1 Signal representation

A typical signal s originates from a photo-detector, described in Chapter 5. It is sampled N times, from t_0 to t_{N-1} every t_{step}. The N samples can be written as:

$$s[t_i]\ t_i \in [t_0..t_{N-1}] \text{ recorded at } t_i = t_0 + i \times t_{step} \tag{6.1}$$

The sampled signal suffers of two types of uncertainty:

1. Sampling time jitter δ_i. The recording of s happens at $t_i + \delta_i$ instead of t_i.
2. Noise ϵ_i. The sampled signal is marred by error and recorded at $s[i] + \epsilon_i$

We consider both errors as uncorrelated, since the noise ϵ_i does not depend, to first order, on the sampling time t_i.

Therefore, if we represent the data taken at the time $t_i + \delta_i$, $s[t_i + \delta_i]$, we are left only with errors on the amplitude of the signal ϵ_i. The true recorded signal is written as:

$$s[t_i + \delta_i] + \epsilon_i\ i \in [0..N-1] \tag{6.2}$$

6.3.2 Time of a signal

The signal representation of the Equation 6.2 can be displayed graphically in Figure 6.1.

The rest of this discussion assumes the initial time of sampling *ref. time* (in blue in Figure 6.1) to be perfectly known, that is defined with respect to a perfect clock. In the sampling window (hatched area in Figure 6.1) the rest of the samples (showed in green) can have sampling jitter $t_i + \delta_i$.

A first method that can be used to determine the time of the signal event is to look at the *fine delay* between the *ref. time* and the extrapolated time of the threshold crossing. This principle is illustrated in Figure 6.1 and will be studied in the following Section.

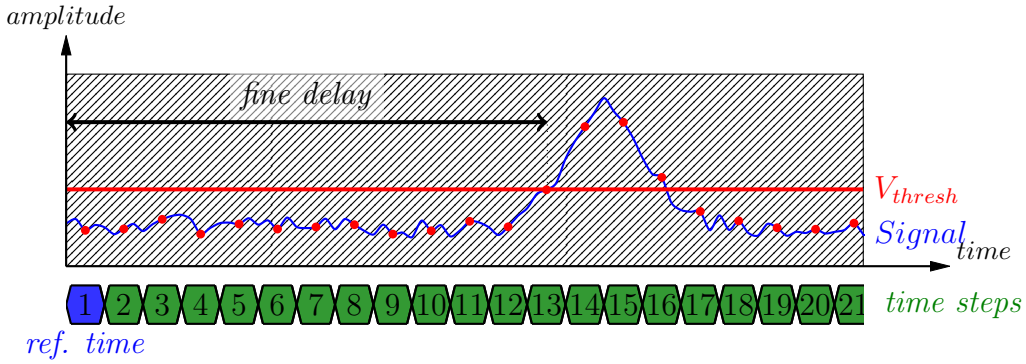


Figure 6.1 – Graphical representation of the continuous signal (blue) and sampled signal (red dots) for each time step (green). The sampling is assumed to be done with respect to a perfectly known clock, providing excellent time resolution with respect to the first sample point *ref. time* (in blue). A first method that can be used to extrapolate the time of the signal event is to look at the *fine delay* between the *ref. time* and the extrapolated time of the threshold crossing.

A second method, which relies on all samples of the signal, will be presented in this Chapter. This method is simply a least mean square algorithm between the sampled signal and its noiseless shape. This method is illustrated in Figure 6.2, where the signal (in blue) is fitted with a template (in black). The *fine delay* extracted is a free parameter of the least mean square fit. In this Chapter, we will show that this method is more precise than using a simple threshold crossing.

To summarize in the rest of the this work the definition of the *Time* of a signal we be taken as:

$$\text{Time} = \textit{ref. time} + \textit{fine time} \tag{6.3}$$

with the *ref. time* a reference time provided by an external, stable clock system, and the *fine time*, the time extracted from the data sample using any kind of time extraction algorithm.

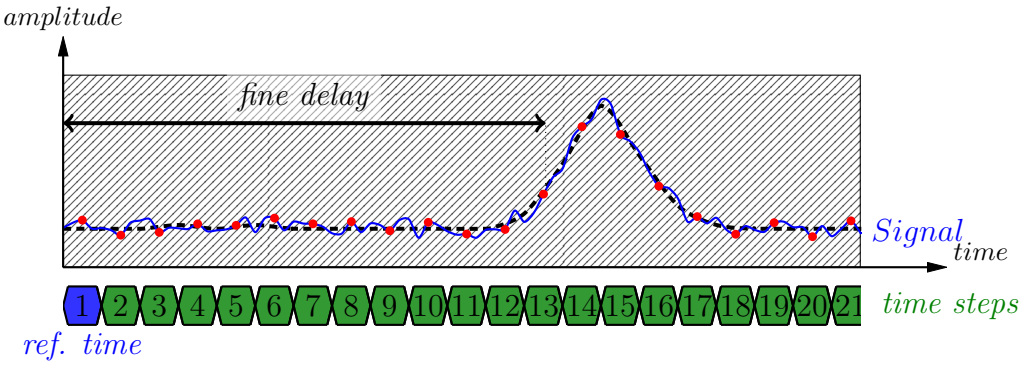


Figure 6.2 – Graphical representation of the continuous signal (blue) and sampled signal (red dots) for each time steps (green). The sampling is assumed to be done with respect to a perfectly known clock, providing excellent time resolution with respect to the first sample point *ref. time* (in blue). A powerful method to extract the precise time of the signal within the sampling window is to use all the recorded points for the time delay estimation by doing a template (black dashed curve) fitting - least mean square.

6.4 Threshold crossing

6.4.1 Introduction

The threshold crossing method is by far the simplest technique to measure time and will be briefly described here, as it permits to introduce the basic concept of precision timing measurement.

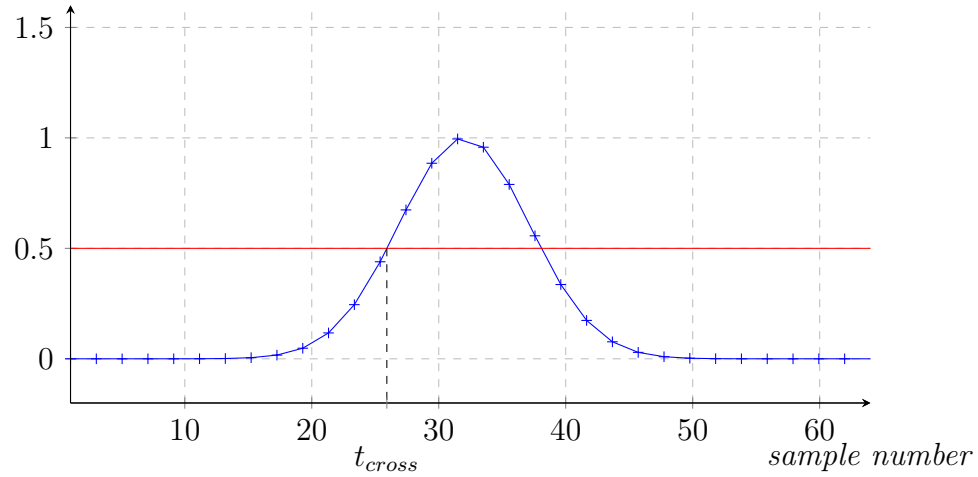


Figure 6.3 – Illustration of the threshold (——) crossing method to find the time of arrival of the pulse (—+—).

This method consist of timing the moment where the signal of interest s will cross a threshold of level a. The time t_{cross} when the signal crosses the threshold is determined with respect to samples i and $i+1$ such as:

$$s[t_i] \leq a \leq s[t_{i+1}] \tag{6.4}$$

In that case, $t_i \leq t_{cross} \leq t_{i+1}$ and the best resolution on t_{cross} is given by $t_{step} = t_{i+1} - t_i$. The resolution of this discrete estimation σ_t is limited by the sampling noise:

$$\sigma_t = \frac{t_{step}}{\sqrt{12}} \tag{6.5}$$

6.4.1.1 Interpolation

In order to get a finer resolution on the crossing time t_{cross} it is possible to do a simple linear interpolation between the samples $s[t_i]$ and $s[t_{i+1}]$. Given a threshold level a falling between the sampled values this allows to continuously extrapolate a crossing time. This simple principle is illustrated in Figure 6.4.

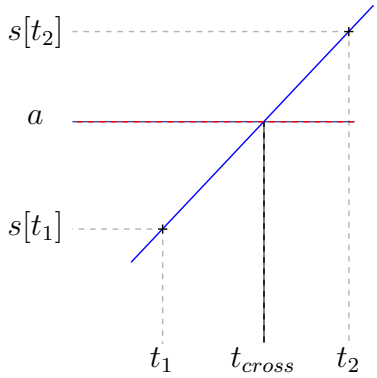


Figure 6.4 – Illustration of the time extraction using linear interpolation between the samples $s[t_i]$ and $s[t_{i+1}]$. The threshold level a is in red and the blue segment is the linear interpolation between the two points. This method allows to extract finer crossing times than t_{step}.

In this case, the signal $s(t)$ between $s[t_i]$ and $s[t_{i+1}]$ is approximated by the linear function $y(t)$:

$$y(t) = \frac{s[t_{i+1}] - s[t_i]}{t_{step}} \times t + \frac{s[t_i]t_{i+1} - s[t_{i+1}]t_i}{t_{step}} \tag{6.6}$$

For any given any threshold level a, the crossing time t_{cross} is determined by:

$$t_{cross} = t_{step}\frac{a - s[t_i]}{s[t_{i+1}] - s[t_i]} + t_i \tag{6.7}$$

6.4.1.2 Performance

The error made by the linear interpolation of a signal s between two points on a curve is bounded by:

$$|s[t] - y(t)| \leq t_{step}^2 \times \frac{1}{8} \max_{y \in [s[t_i], s[t_{i+1}]]} |s''(t)| \tag{6.8}$$

This error can be reduced if necessary using higher order polynomial interpolations (2^{nd} order, 3^{rd} order, ...) or using spline interpolation, depending on the signal shape.

But if the threshold a is set on a region where the signal is linearl ($s''(t) \approx 0$) at the point of crossing, the error made by the linear interpolation is minimum. When using the threshold method it is therefore important to set the threshold level such as it intercepts the pulse of interest in its most linear region.

The performance of the threshold crossing method is in fact mostly limited by the noise on the sampled signal. Indeed the true sampled signal, including its noise model from Equation 6.2, can be written as follow:

$$s[t_i + \delta_i] + \epsilon_i \approx s[t_i] + \delta_i s'[i] + \epsilon_i \tag{6.9}$$

For the purpose of the error computation on the crossing time t_{cross}, due to the noise on $s[t_i]$ and $s[t_{i+1}]$, the noise on both samples is reduced to the term n_1 and n_2 such as, with $s[t_i]^*$ is the noiseless expression of $s[t_i]$ we have:

$$\begin{aligned} s[t_i] &= s[t_i]^* + n_1 \\ s[t_{i+1}] &= s[t_{i+1}]^* + n_2 \end{aligned} \tag{6.10}$$

The error ϵ on the crossing time t_{cross} due to the noise terms n_1 and n_2 is expressing in Equation 6.11. Computation of the error term is omitted and the result displayed below:

$$\begin{aligned} \epsilon =& t_{step}\frac{a - s[t_i]^*}{s[t_{i+1}]^* - s[t_i]^*} + t_i - t_{step}\frac{a - s[t_i]}{s[t_{i+1}] - s[t_i]} + t_i \\ =& t_{step}\left(\frac{a - s[t_i]^*}{s[t_{i+1}]^* - s[t_i]^*} - \frac{a - s[t_i]^* - n_1}{s[t_{i+1}]^* - s[t_i]^* + n_2 - n_1}\right) \\ =& \dots \\ =& t_{step}\frac{n_1(s[t_{i+1}]^* - a) - n_2(s[t_i]^* - a)}{s[t_{i+1}]^* - s[t_i]^*} \end{aligned} \tag{6.11}$$

This result is quite remarkable and best understood when looking at the variance of the error ϵ. If we assume a uniform, uncorrelated noise for all samples $\mathrm{Var}(n_i) = \sigma_n^2$ we can write:

$$\mathrm{Var}(\epsilon) = \left(\sigma_n t_{step}\frac{\sqrt{(s[t_{i+1}]^* - a)^2 + (s[t_i]^* - a)^2}}{s[t_{i+1}]^* - s[t_i]^*}\right)^2 \tag{6.12}$$

The representation of the variance versus the crossing threshold a is represented in Figure 6.5. We can observe that, if the signal *grows linearly* in the region of the crossing ($s''(t) \approx 0$), then the time resolution reachable is much finer than with the previous coarse approximation $t_{cross} \sim t_{step}$. Here the resolution is basically set by the noise level σ_n of the signal.

We observe that the interpolation method is more efficient when the threshold level a is set between sample $s[t_i]$ and $s[t_{i+1}]$. In that case we gain statistically on the noise by a factor $\sqrt{2}$. However, this condition cannot be controlled experimentally and we have to assume that the best precision of this method is given by $t_{step} \times \sigma_n$. To obtain better performances more points should be considered when doing the interpolation on the signal s. This option will be presented in a following Section.

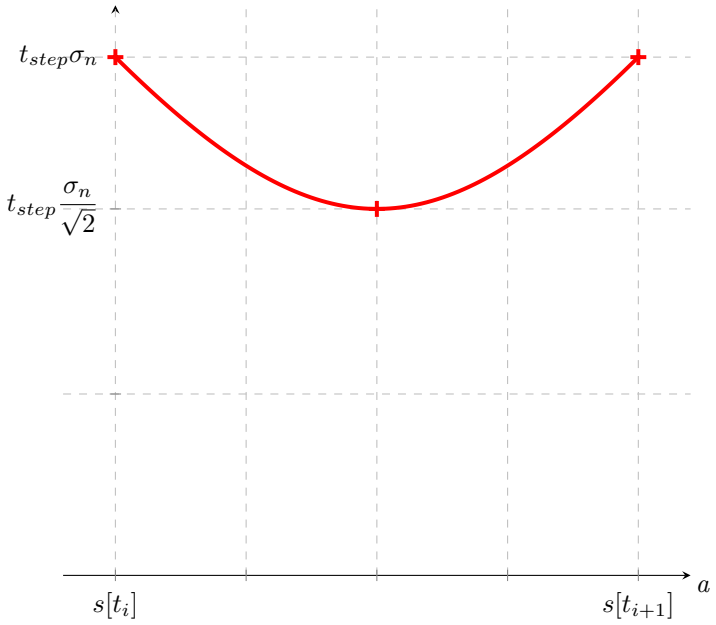


Figure 6.5 – Variance of the noise $\mathrm{Var}(\epsilon)$ obtained using the threshold crossing technique in the interval $s[t_i], s[t_{i+1}]$. We can observe that when the threshold a falls near the sample $s[t_i]$ or $s[t_{i+1}]$ the linear interpolation does not help reduce the noise. In between the sample the noise is reduced by a factor $\sqrt{2}$. This results holds only in the case of Equation 6.8.

6.4.1.3 Amplitude variation

The effect of signal amplitude variation on the time resolution using the threshold method is illustrated in Figure 6.6. We can see that any variation in the signal amplitude A leads to large variations on the time t_i extracted with this method: this is the time walk phenomenon.

A simple way of suppressing f the amplitude variation effect is to fix the threshold level at a constant fraction of the amplitude of the signal. This is done using the Constant Fraction Discriminator method (CFD). Therefore we can expect this method - in its crude implementation, with linear signal interpolation - to give the time resolution obtained in Equation 6.12.

6.4.1.4 Baseline shift

Another big disadvantage of the threshold method is its sensitivity to baseline shifts. In the previous study the signal reference $s[0]$ was supposed to be constant. In reality the entire waveform $s[t_i]$ $\forall t_i \in [t_0..t_{N-1}]$ is frequently affected by low-frequency noise shifting the entire base of the signal.

The effect of baseline shift can be seen in Figure 6.7, where a large number of baselines have been recorded during an experiment and are displayed. Some of the waveforms are shifted downwards compared to others. The threshold method cannot accommodate for such shifts and will produce large errors.

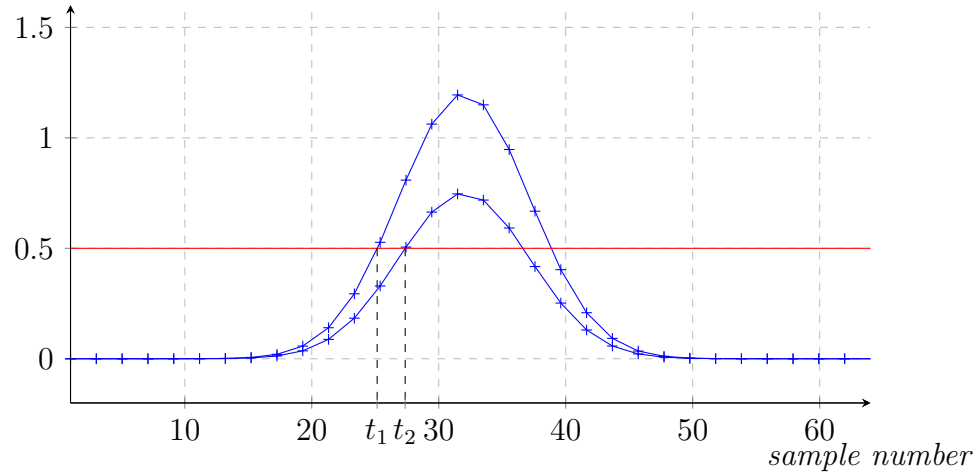


Figure 6.6 – Illustration of the effect of amplitude variation on the signal (—+—) for a given threshold level (——).

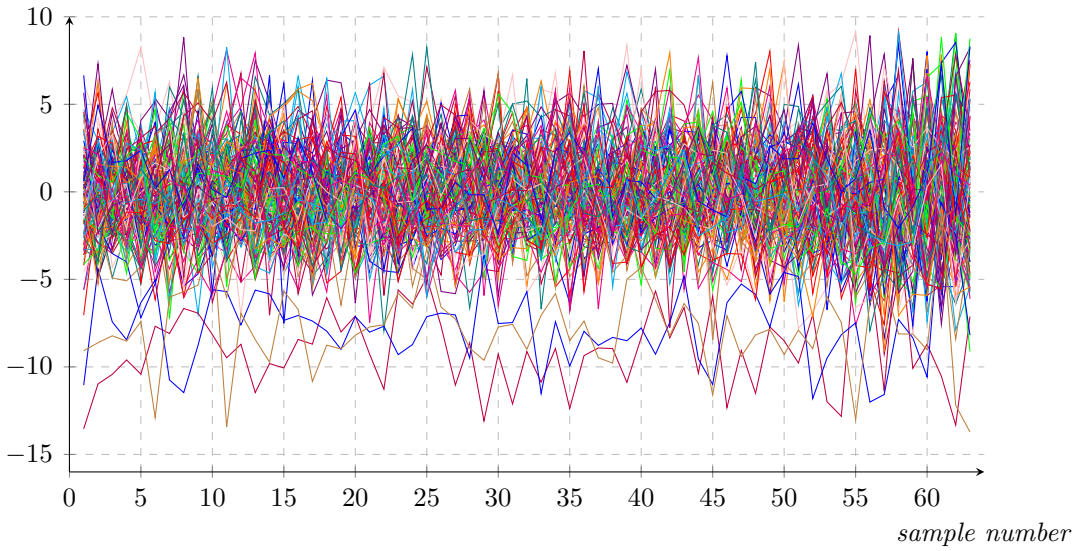


Figure 6.7 – Baseline shift illustration while recording noise.

6.4.2 Conclusion

In conclusion, for fine time measurements, the threshold method cannot be used in its simplest version due to its dependence on baseline shifts and amplitude variations. These effects can be compensated at the cost of a higher number of computation steps using higher order signal interpolation, baseline subtraction and constant fraction discrimination. However, in some cases, where coarse time resolutions are sufficient the threshold method can be implemented in its simple form.

6.5 Cross-correlation

6.5.1 Introduction

The cross-correlation measures the similarity between two waveforms versus a time-lag applied to one of them. It can be used in our particular case to find the particular time lag t_{lag} between our recorded signal s of unknown time and a reference template t of perfectly known time. The template t is made similar to the sampled signal in order to maximize the similarity between the two curves and to find the time lag t_{lag} more precisely. For a signal s sampled on N points, the cross correlation between s and the template t is written as follows:

$$c[t_k] = \sum_{i=0}^{N-1} s[t_i] \times t[t_{i+k}] \; k \in [-N+1..N-1] \tag{6.13}$$

The cross-correlation $c[t_k]$ takes its maximum value when $s[t_i] = t[t_{i+k}]$:

$$t_k = t_{lag} \tag{6.14}$$

Therefore, finding where the cross-correlation takes its maximum value precisely is the key to measure the time-difference between the sample signal and the reference, and we can see that since all the sampled points $s[t_i]$ are used to compute the cross-correlation, we have a higher statistics than the threshold method and results should be more precise.

$$t_{lag} = t_k \left| c[t_k] = \max_{k \in [-N+1, N-1]} \sum_{0}^{N-1} s[t_i] t[t_{i+k}] \right. \tag{6.15}$$

6.5.1.1 Baseline shifts

Compared to the threshold crossing, the cross-correlation does not suffer from baseline shifts. Indeed if we have a signal s affected by a base line shift of magnitude B, the maximum of cross-correlation is written as:

$$\begin{aligned} t_{lag} &= t_k \left| c[t_k] = \max_{k \in [-N+1, N-1]} \sum_{0}^{N-1} \left[s[t_i] + B \right] t[t_{i+k}] \right. \\ &= t_k \left| c[t_k] = \max_{k \in [-N+1, N-1]} \sum_{0}^{N-1} s[t_i] t[t_{i+k}] \right. \end{aligned} \tag{6.16}$$

since the sum $\sum_0^{N-1} B[t_{i+k}]$ is a constant. The baseline shift will only create an offset for the whole correlation function c.

This technique is only sensitive to baseline fluctuations resulting in a non-uniform behavior over the whole range of samples. These fluctuations can be considered as oscillations with a frequency f_N. Baseline fluctuations with oscillation times much lower than the total recording depth $N \times t_{step}$ will be seen as a continuous signal

(baseline shift) and will not affect the cross-correlation algorithm. If we have recorded N points with a step size t_{step} we will not be sensitive to frequencies f_N such as:

$$f_N \ll \frac{1}{N t_{step}} \tag{6.17}$$

Reducing the number of points sampled and increasing the sampling frequency $f_s = {}^1/_{t_{step}}$ makes us more independent of low-frequency noise, cutting it off. However this comes at the cost of the need for faster signals and higher bandwidth; sufficient to record the signal of interest on a shorter depth. The minimum frequency f_{min} for a signal to be fully recorded on N points spaced by t_{step} being:

$$f_{min} = \frac{1}{N t_{step}} \tag{6.18}$$

Table 6.1 – Upper bound on the baseline cutoff frequency f_N for various sampling frequencies and number of samples. A factor of 100 is applied (*i.e.* $f_N = f_{min}/100$).

Sampling frequency	Number of points	Baseline cutoff frequency f_N	Minimum signal frequency f_{min}
3 GSa/s	1024	30 kHz	3 MHz
10 GSa/s	256	400 kHz	40 MHz
10 GSa/s	64	1.5 MHz	150 MHz

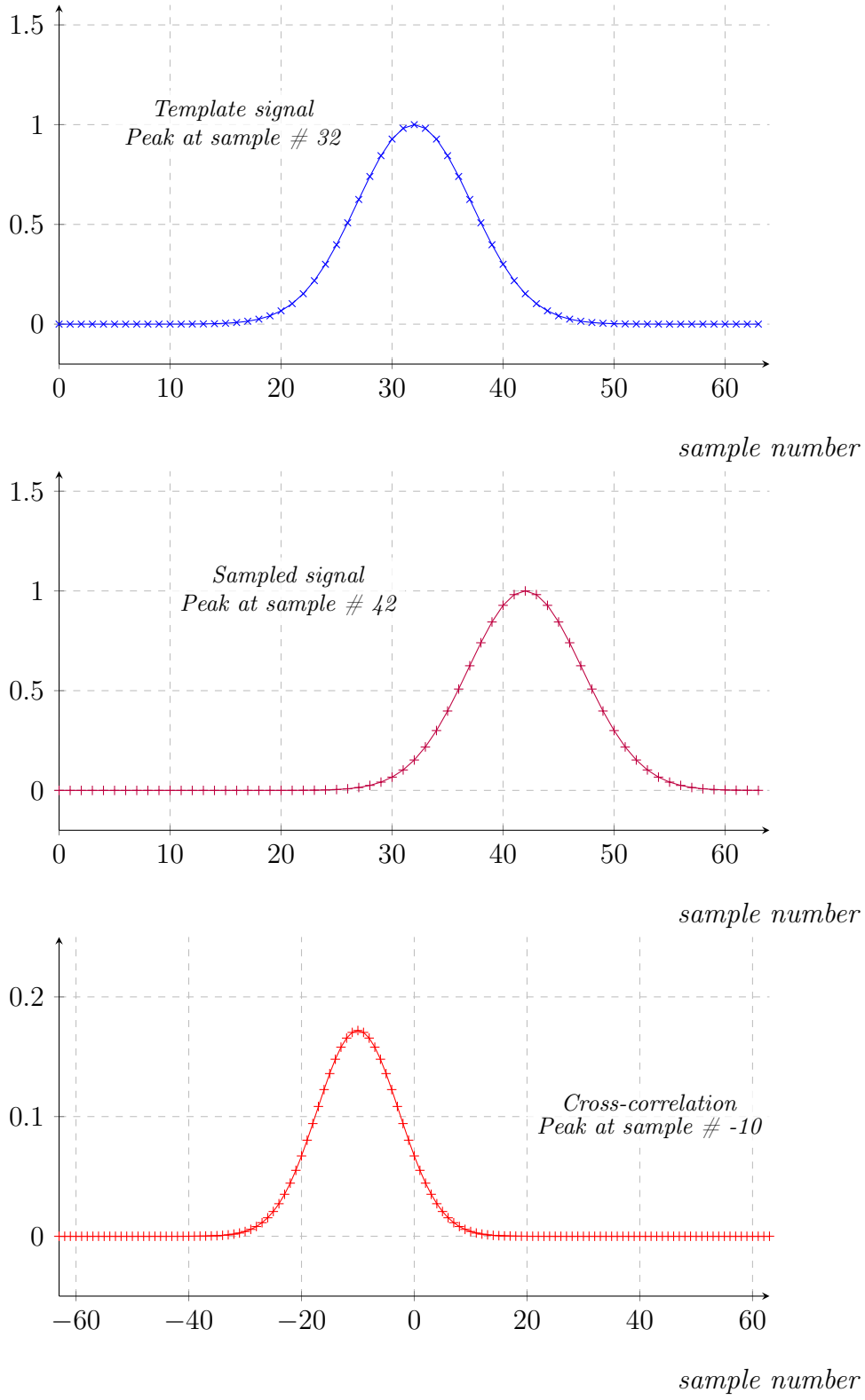


Figure 6.8 – Example of cross-correlation between a Gaussian pulse sampled on 64 points (—×—) with a peak at sample number 32 and a template signal (—+—) with a peak at sample number 42. The location of the maximum of the cross-correlation (—+—) gives the delay t_{lag} between the pulses, here $t_{lag} = 10$.

6.5.2 Template signal creation

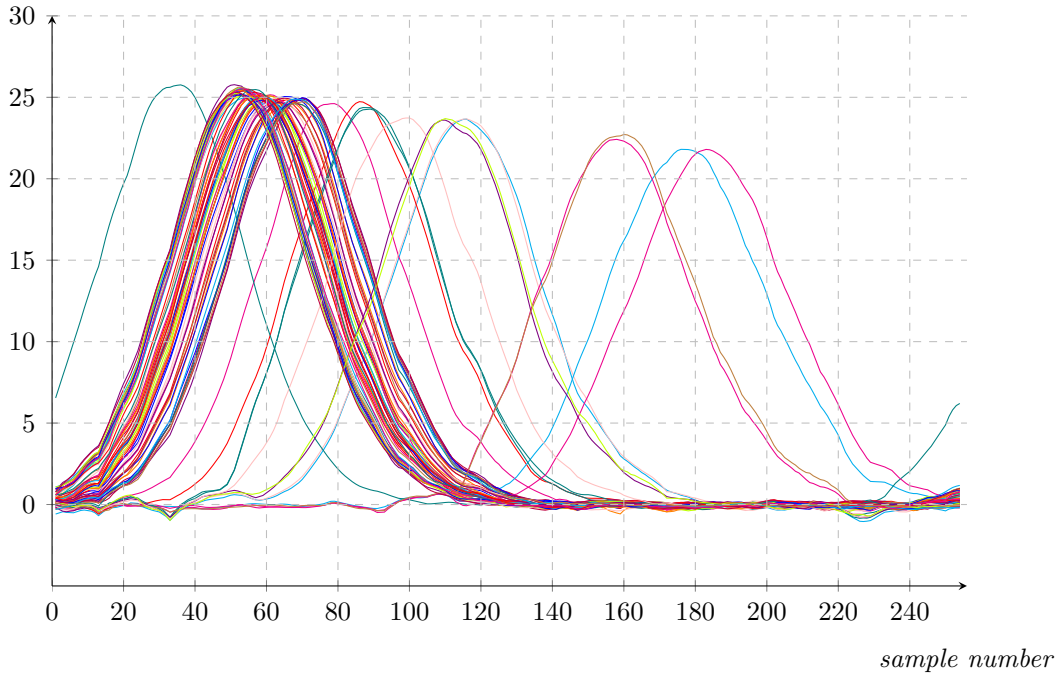


Figure 6.9 – Sampled pulses used to create the signal template.

The template signal is the waveform used as a reference to extract the time of all sampled signals. In order to be able to make a precise time delay measurement we saw that it is essential to know it with a timing structure better than the signal.

The template signal must reflect the shape of the signal to be measured. At the maximum of correlation between the template and the signal we expect to have $t[i+k] \approx s[i]$. In order to do so, the template is built by averaging a large number of sampled signals, effectively removing noise contributions.

We have seen that the tail of the signal from a photo-detector can have a much larger variance than its rising edge (due to bandwidth limitation effects for example). The time information contained in the rising edge is therefore much more important than in the tail. To reflect this effect, the template signal is normalized to the standard deviation of the signal σ_t for every time step. σ_t is also calculated using a high number of sampled signal.

If we use M input signals to build the template $t[k]$:

$$t[t_k] = \sum_{i=0}^{M-1} \frac{s_i[t_k]}{\sigma_t} \tag{6.19}$$

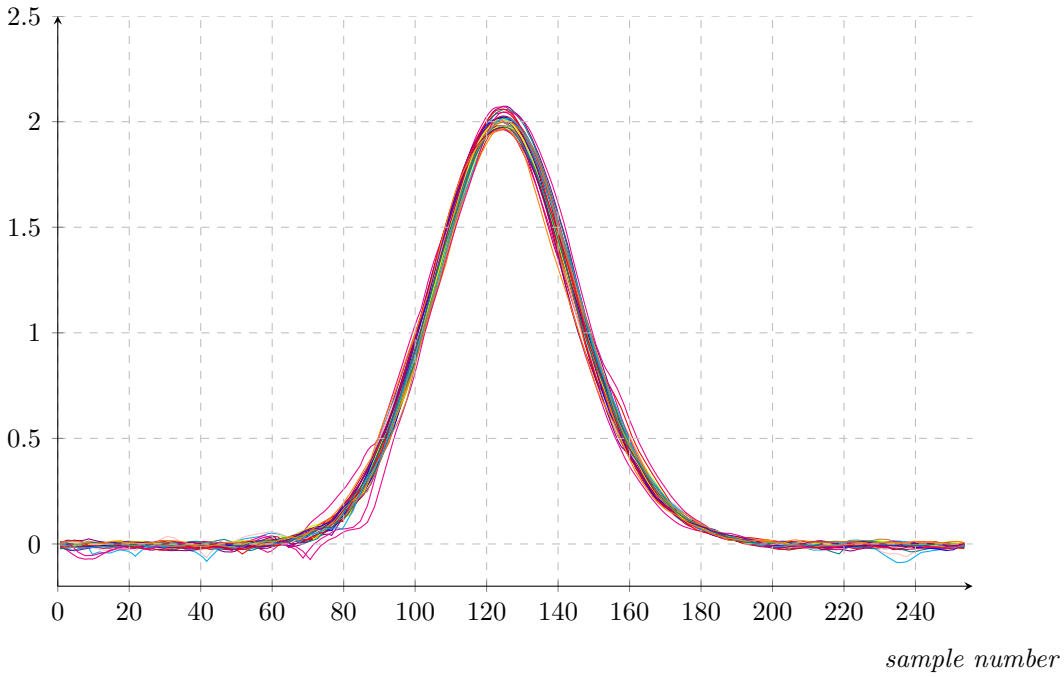


Figure 6.10 – Overlaid sample pulses used to create the template.

6.5.3 Time delay estimation precision

Without interpolation, the best resolution on the cross-correlation maximum is set by the step t_{step} of the signal sampled. The template t, however, could be known with a much finer sampling step: $\frac{t_{step}}{10}$ or $\frac{t_{step}}{100}$. Practically, this is equivalent to using a finer template t and padding the signal s with zeros such that the cross-correlation c of Equation 6.13 becomes :

$$c[t_k] = \sum_{i=0}^{N-1} s[t_i] \times t[t_{i+k/100}] \; k \in [-100N+1, 100N-1] \tag{6.20}$$

With that precision on the template, the peak location of the cross-correlation can be then calculated with a precision of $\frac{t_{step}}{10}$ or $\frac{t_{step}}{100}$.

Practically this is equivalent to use a finer template and the signal padded with zeros where:

$$s[t_i] \neq s[t_{step} \times i] \; i \in [0, 100N-1] \tag{6.21}$$

This principle is illustrated in Figure 6.11 where the sampled signal is padded with zeros in between samples.

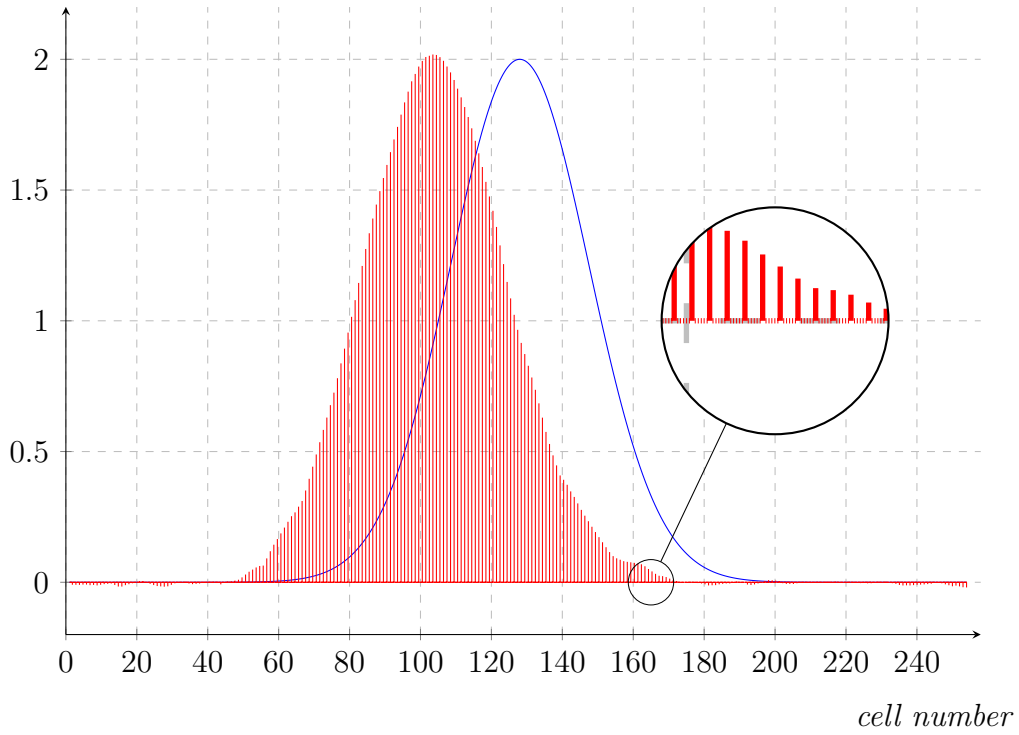


Figure 6.11 – Zero-padding of the sampled data (· ·), to increase the resolution of the cross-correlation algorithm of the template (——). Zero-padding with m zeros, increases the number of operations by a factor m.

6.5.4 Precise maximum determination

We can note that the cross-correlation presents a symmetric peak around $t_m = t_{lag}$. The peak location can be computed even more precisely by approximating c by a convex parabola in the neighborhood of its maximum.

$$c[t] = at^2 + bt + c \tag{6.22}$$

Here a, b and c are the parameters of the correlation fit. Using Equation 6.22, it is straightforward to calculate a more precise value of the maximum location of the correlation t_m.

$$t_m = -\frac{b}{2a} \tag{6.23}$$

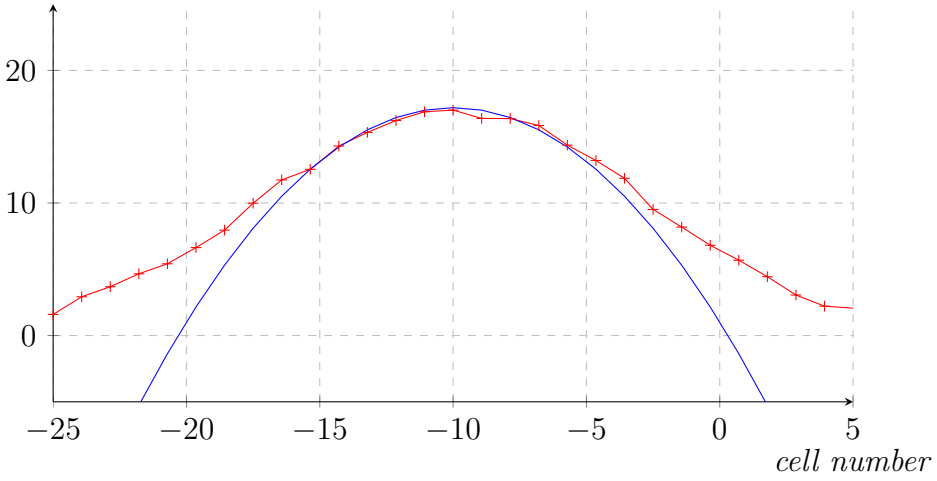


Figure 6.12 – Zoomed in view of the parabolic fit (—) of a noisy Gaussian pulse (—+—) near the extremum. Interpolation allows a more accurate estimate of the extremum of the cross-correlation, rather than using the position of the maximum.

6.5.5 Signal-to-noise ratio

The parabola approximation of the cross-correlation gives a simple way to estimate the time delay t_{lag} between the template signal t and the real signal s. The maximum of cross-correlation is obtained when the sampled signal best matches the template signal: $t[t_{i+k}] \approx s[t_i]\ \forall i$. It corresponds to:

$$t_k = t_{lag} \tag{6.24}$$

In this case we can approximate the template near the maximum using a Taylor series expansion:

$$t\left[t_{lag} + \delta_t\right] \approx s[t_i] + \delta_t s'[t_i] + \frac{\delta_t^2}{2} s''[t_i] \text{ with } \delta_t \ll t_{step} \tag{6.25}$$

Therefore, we can write the cross-correlation close to the maximum as:

$$c\left[t_{lag}+\delta_t\right] \approx \sum_0^{N-1} s[t_i]^2 + \delta_t \sum_0^{N-1} s[t_i]s'[t_i] + \delta_t^2 \sum_0^{N-1} \frac{s[t_i]s''[t_i]}{2} = a\delta_t^2 + b\delta_t + c \quad (6.26)$$

Using this expression, we verify that the cross-correlation can be approximated by a parabola close to its extremum. Around the extremum we have $b = 0$:

$$\sum_0^{N-1} s[t_i]s'[t_i] = 0 \quad (6.27)$$

A consequence of the above formula is that:

$$\left(\sum_0^{N-1} s[t_i]s'[t_i]\right)' = 0 \quad (6.28)$$

$$\Leftrightarrow \quad \sum_0^{N-1} s[t_i]s''[t_i] = -\sum_0^{N-1} s'[t_i]^2 \quad (6.29)$$

We verify by this formula that $\sum_0^{N-1} s[t_i]s''[t_i] \leq 0$, so we have a maximum of correlation at t_{lag} (again this is expected for a maximum).

Thus, the cross-correlation near the maximum value is given by the formula:

$$c\left[t_{lag}+\delta_t\right] \approx \sum_0^{N-1} s[t_i]^2 - \delta_t^2 \sum_0^{N-1} \frac{s'[t_i]^2}{2} \quad (6.30)$$

where the term $\sum_0^{N-1} s[i]^2$ corresponding to the maximum of correlation but also to the energy E of the signal s.

The signal however is affected by the noise n. A more in depth study of the source and effect of this noise will be carried in the following Section. Basically the effect of this noise it to shift randomly the extremum of the cross-correlation function. This has for effect to widen the resolution of the maximum location and worsen the time resolution. This effect is shown in Figure 6.13.

The effect of the noise is to miss-estimate the maximum of the cross-correlation at $t = t_{lag}$. Instead, due to the noise, the maximum will be shifted by a random amount t_m, yielding to an uncertainty on the precise time of the signal s.

6.5.6 Signal with noise

With both the sampling time jitter and the noise on $s[i]$ we write the cross-correlation as:

$$c[t_k] = \sum_0^{N-1} t[t_{i+k}]\left(s[t_i+\delta_i]+\epsilon_i\right) \quad (6.31)$$

Assuming that $\delta_i << t_{step}$, we can develop s at the first order and write:

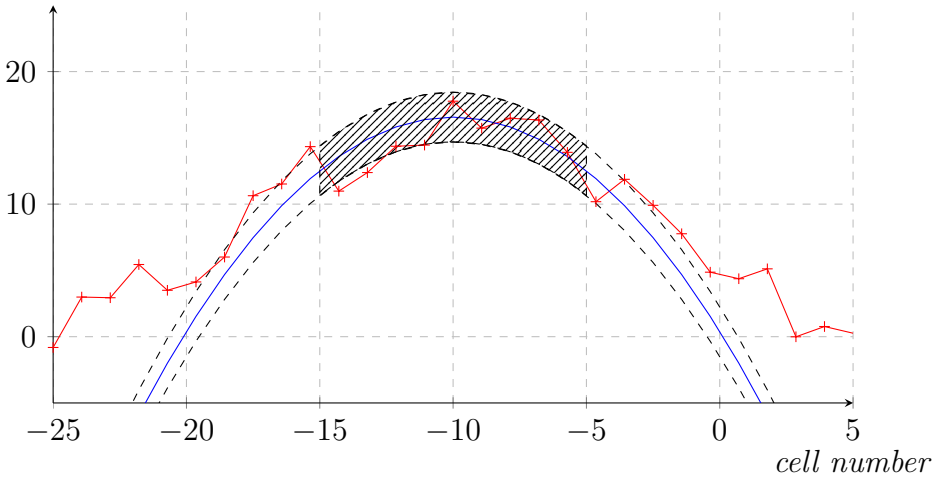


Figure 6.13 – Parabolic fit (——) of a noisy Gaussian pulse (—+—). The noise amplitude on the cross-correlation plot define a region (in gray) and thus there is an uncertainty on the extremum location.

$$s[t_i + \delta_i] \approx s[i] + \delta_i s'[i] \tag{6.32}$$

Equation 6.31 can be expanded into three sums:

$$c[t_k] = \sum_{0}^{N-1} s[t_i]t[t_{i+k}] + \sum_{0}^{N-1} \delta_i s'[t_i]t[t_{i+k}] + \sum_{0}^{N-1} \epsilon_i t[t_{i+k}] \tag{6.33}$$

At the maximum for c, we use the Taylor series approximation of the template t presented in Equation 6.25:

$$\begin{aligned} c[t_{lag} + \delta_t] \approx & \sum_{0}^{N-1} s[t_i]\left(s[t_i] + \delta_t s'[t_i] + \frac{\delta_t^2}{2}s''[t_i]\right) \\ & + \sum_{0}^{N-1} \delta_i s'[t_i]\left(s[t_i] + \delta_t s'[t_i] + \frac{\delta_t^2}{2}s''[t_i]\right) \\ & + \sum_{0}^{N-1} \epsilon_i \left(s[t_i] + \delta_t s'[t_i] + \frac{\delta_t^2}{2}s''[t_i]\right) \end{aligned} \tag{6.34}$$

$$\begin{aligned} \approx & \sum_{0}^{N-1} s[t_i]^2 + \delta_i s[t_i]s'[t_i] + \epsilon_i s[t_i] \\ & + \delta_t \sum_{0}^{N-1} s[t_i]s'[t_i] + \delta_i s'[t_i]^2 + \epsilon_i s'[t_i] \\ & + \delta_t^2 \sum_{0}^{N-1} \frac{s[t_i]s''[t_i]}{2} + \delta_i \frac{s'[t_i]s''[t_i]}{2} + \epsilon_i \frac{s''[t_i]}{2} \end{aligned} \tag{6.35}$$

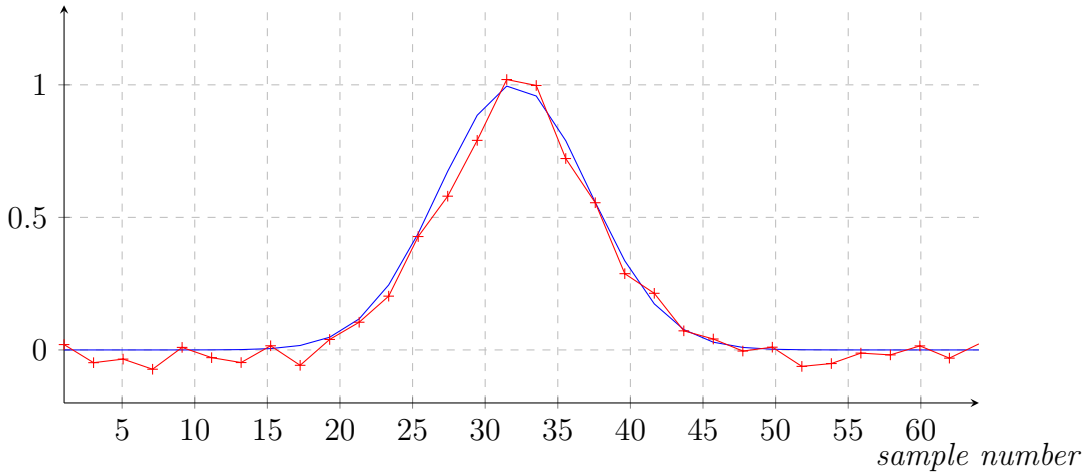


Figure 6.14 – Illustration of a real signal affected by sampling jitter and noise (—+—), compared to the ideal noiseless signal (——).

In the second part of Equation 6.35, we recognize the parabolic formulae $a\delta_t^2 + b\delta_t + c$. And if we write t_m, the new extremum of the parabola with noise, is expressed as:

$$\begin{aligned} t_m &= \frac{-b}{2a} \\ &= -\frac{\sum_0^{N-1} \delta_i s'[t_i]^2 + \epsilon_i s'[t_i]}{\sum_0^{N-1} s[t_i]s''[t_i] + \delta_i s'[t_i]s''[t_i] + \epsilon_i s''[t_i]} \\ &\approx \frac{\sum_0^{N-1} \delta_i s'[t_i]^2 + \epsilon_i s'[t_i]}{\sum_0^{N-1} s'[t_i]^2} \end{aligned} \tag{6.36}$$

We verify that the variation of the extremum location t_m is purely due to noise on the signal. If $\delta_i = 0$ and $\epsilon_i = 0$ we have $t_m = 0$. The contribution of the sampling time jitter δ_i and of the noise ϵ_i produces two distinct error terms.

$$t_m = \frac{\sum_0^{N-1} \delta_i s'[t_i]^2}{\sum_0^{N-1} s'[t_i]^2} + \frac{\sum_0^{N-1} \epsilon_i s'[t_i]}{\sum_0^{N-1} s'[t_i]^2} \tag{6.37}$$

In the following Section the source of the two distinct noise terms of Equation 6.37 are investigated and an evaluation of their magnitude in terms of physical quantities such as the pulse amplitude A, number of points n or sampling time step t_{step} is given.

6.5.7 Effect of sampling time jitter

The effect of the sampling jitter is to randomly offset the sampling times t_i in the x-axis direction, as shown in Figure 6.15.

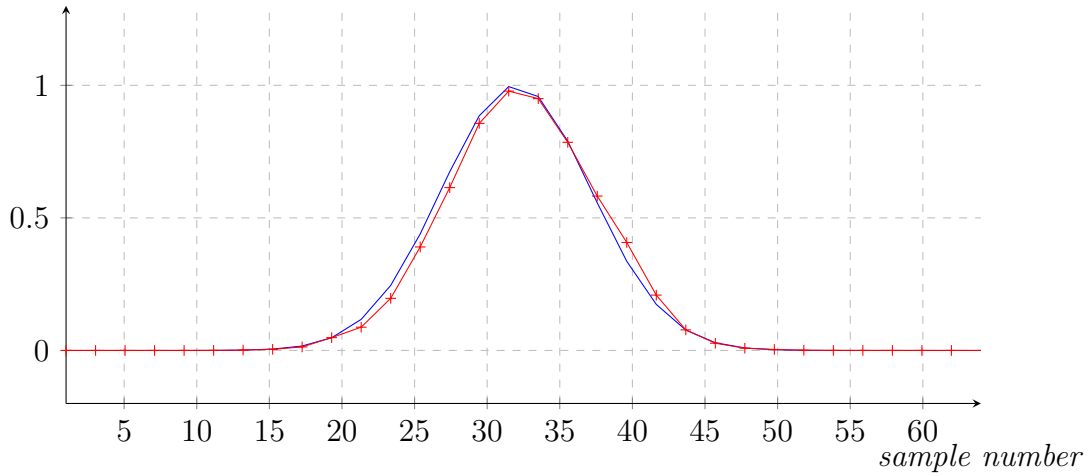


Figure 6.15 – Illustration of the effect of sampling jitter on the pulse shape (—+—). The effect of the sampling jitter is to randomly offset the samples in time, as can be seen comparing it to the noiseless signal (——).

The error term n_j on t_m due to the sampling noise jitter δ_i is written (from Equation 6.37):

$$n_j = \frac{\sum_0^{N-1} \delta_i s'[t_i]^2}{\sum_0^{N-1} s'[t_i]^2} \tag{6.38}$$

The error n_j due to the sampling time jitter is the weighted sum of the jitter on the samples δ_i by the weights $w_i = \frac{s'[t_i]^2}{\sum_0^{N-1} s'[t_i]^2}$. These weights w_i are normalized, such that:

$$\sum_0^{N-1} w_i = 1 \tag{6.39}$$

If we suppose that the sampling time jitter noise δ_i of all the samples are uncorrelated, and of same variance $\mathrm{Var}(\delta_i) = \sigma_i^2$, the variance of the error n_j is given by:

$$\mathrm{Var}(n_j) = \sigma_j^2 \sum_0^{N-1} w_i^2 \tag{6.40}$$

Then, the variance attains its minimum value when all weights are equal and its maximum value when all but one weight are zero. However in the sum of the N samples $s'[t_i]^2$ only the n non-zero samples of the signal s (see Figure 6.17) contribute to the sampling noise jitter and we have:

$$\frac{\sigma_j^2}{n} \leq \mathrm{Var}(n_j) \leq \sigma_j^2 \tag{6.41}$$

It is interesting to note that the variance of the noise due to the sampling time jitter does not depend, to the first order, on the signal amplitude.

- $\boldsymbol{\sqrt{\mathbf{Var}(n_j)} \propto \frac{1}{\sqrt{n}}}$. In case of sampling time jitter, a larger number of samples on the signal will improve the precision on the time delay estimation statistically.
- $\boldsymbol{\sqrt{\mathbf{Var}(n_j)}}$ does not depend on the sampling time step, in case of sampling time jitter. [1]

For the typical signals that we are considering, such as the Gaussian pulse of Figure 6.15, we can make the coarse approximation of a triangular shape where $s' \approx$ cst. With that approximation the weights w_i are uniformly distributed and:

$$\boldsymbol{\mathrm{Var}(n_j) \approx \frac{\sigma_j^2}{n}} \tag{6.42}$$

[1] In reality the standard deviation σ_j of the sampling time jitter is usually a function of the time step t_{step}: it is easier to have 1ps sampling jitter at 10Gs.s^{-1} than at 1Ms.s^{-1}.

6.5.8 Effect of noise in the sampled signal

The effect of the true noise ϵ_i is to randomly offset the samples $s[t_i]$ following the y-axis direction. See Figure 6.16.

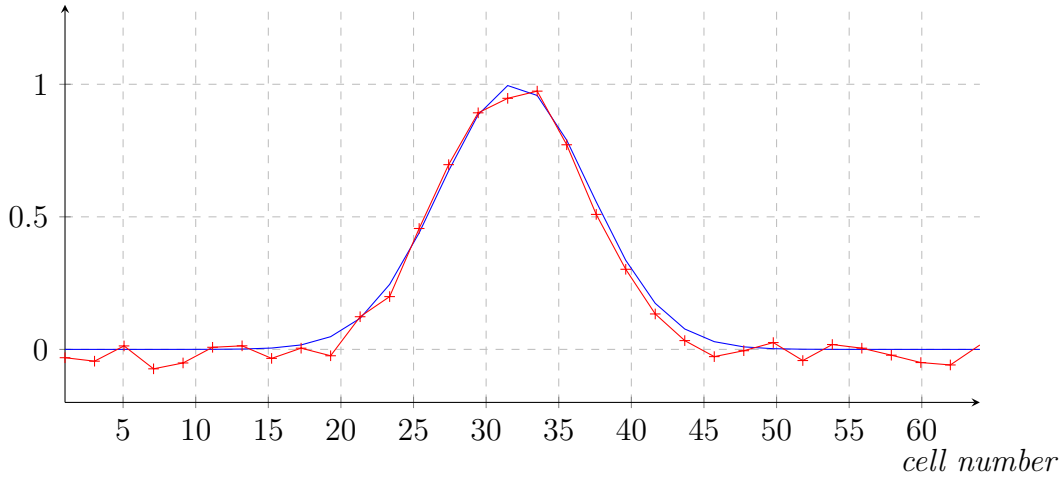


Figure 6.16 – Illustration of the effect of noise on the pulse shape (—+—). The effect of the signal noise is to randomly offset the signal in the y-axis direction and be seen by comparing it to the true, noiseless signal(——).

The error term n_ϵ on t_m due to the signal noise ϵ_i is written (from Equation 6.37):

$$n_\epsilon = \frac{\sum_0^{N-1} \epsilon_i s'[t_i]}{\sum_0^{N-1} s'[t_i]^2} \tag{6.43}$$

If we suppose the noise ϵ_i on all the samples are uncorrelated and of same variance $\mathrm{Var}(\epsilon_i) = \sigma_n^2$, the variance of the error n_ϵ is given by:

$$\mathrm{Var}(n_\epsilon) = \sigma_n^2 \frac{1}{\sum_0^{N-1} s'[t_i]^2} \tag{6.44}$$

Using the Taylor expansion between samples $s[t_i]$ and $s[t_{i+1}]$, we can write:

$$s'[t_i] \approx \frac{s[t_{i+1}] - s[t_i]}{t_{step}} \tag{6.45}$$

By summing Equation 6.45 from t_0 to t_{lag} and from t_{N-1} to t_{lag} we have:

$$\begin{aligned} \sum_0^{N-1} |s'[t_i]| &\approx \sum_0^{lag} |s'[t_i]| + \sum_{N-1}^{lag} |s'[t_i]| \\ &\approx \sum_0^{lag} \frac{s[t_{i+1}] - s[t_i]}{t_{step}} + \sum_{N-1}^{lag} \frac{s[t_i] - s[t_{i+1}]}{t_{step}} \\ &\approx \frac{2A}{t_{step}} \end{aligned} \tag{6.46}$$

with A the maximum amplitude taken by s at $t = t_{lag}$. And using the Cauchy-Schwarz inequality we have:

$$\left|\sum_{0}^{N-1}|s'[t_i]|\right|^2 \leq \sum_{0}^{N-1} s'[t_i]^2 \sum_{0}^{N-1} 1 \tag{6.47}$$

The sum can be reduced to the n points where the signal s is non-zero, and the lower bound on the sum $\sum_0^{N-1} s'[t_i]^2$ is:

$$\frac{4A^2}{nt_{step}^2} \leq \sum_{0}^{N-1} s'[t_i]^2 \tag{6.48}$$

The lower bound is actually reached when $s' \approx cst$. With this bound and the Equation 6.44, calculated previously, we set the lower bound on the variance on the error n_ϵ by:

$$\frac{nt_{step}^2\sigma_n^2}{4A^2} \leq \mathrm{Var}(n_\epsilon) \tag{6.49}$$

The following observations can be made about Equation 6.49:

- $\boldsymbol{\sqrt{\mathbf{Var}(n_\epsilon)} \propto \sqrt{n}}$. In case of noise on the signal, a larger number of samples on the signal will degrade the time delay estimation. This effects counters the opposite effect in case of sampling time jitter.[2]
- $\boldsymbol{\sqrt{\mathbf{Var}(n_\epsilon)} \propto \frac{\sigma_n}{A}}$. In case of noise on the signal, a signal with a higher signal to noise ratio will improve the precision on the time delay estimation.
- $\boldsymbol{\sqrt{\mathbf{Var}(n_\epsilon)} \propto t_{step}}$. In case of noise on the signal, a faster sampling will improve the precision on the time delay estimation.

For the typical signals we are considering, such as a Gaussian pulse of Figure 6.16, we can make the coarse approximation of a triangular shape where $s' \approx$ cst. With that approximation the sum $\sum_0^{N-1} s'[t_i]^2$ reaches its lower bound value $\frac{4A^2}{nt_{step}^2}$, so we can write:

$$\mathbf{Var}(n_\epsilon) \approx \frac{\boldsymbol{nt_{step}^2\sigma_n^2}}{\boldsymbol{4A^2}} \tag{6.50}$$

6.5.9 Performances comparison and discussion

The variance of the total noise $n_t = n_j + n_\epsilon$ (from jitter and sampling noise) on the extracted time t_m by the cross-correlation algorithm can be written as:

$$\mathbf{Var}(n_t) \sim \frac{\boldsymbol{nt_{step}^2\sigma_n^2}}{\boldsymbol{4A^2}} + \frac{\boldsymbol{\sigma_j^2}}{\boldsymbol{n}} \tag{6.51}$$

[2] As strange as it may appear, we are here considering a higher number of points on a signal for a given t_{step} this means that we are in reality sampling a signal with a lower bandwidth.

It is important to note that this shows that in order to obtain good time resolution not only the sampling noise or jitter has to be reduced, but both simultaneously. Good timing cannot be obtained with poor signal-to-noise ratio nor poor sampling time jitter.

In the following Sections we are going to simulate the effect of the various parameters, amplitude A, number of points n, sampling noise σ_n and jitter σ_j, in order to verify the result of Equation 6.51.

6.6 Simulation of the noise model

In this Section we present a MATLAB [1] simulation performed in order to verify the validity of the results computed in the previous Section. Equation 6.51 in particular is simulated.

This simulation consists of generating a noisy signal and extracting its time using the cross-correlation method presented. Several (100) consecutive time measurements are performed and the *variance* of the 100 measurments is computed.

In order to verify the dependence upon the parameters of Equation 6.51 several simulations are performed, in each one of them one of the following parameters is varied (see illustration in Figure 6.17):

- A, the amplitude of the signal.
- t_{step}, the sampling time step.
- n, the number of non-zero points on the signal.

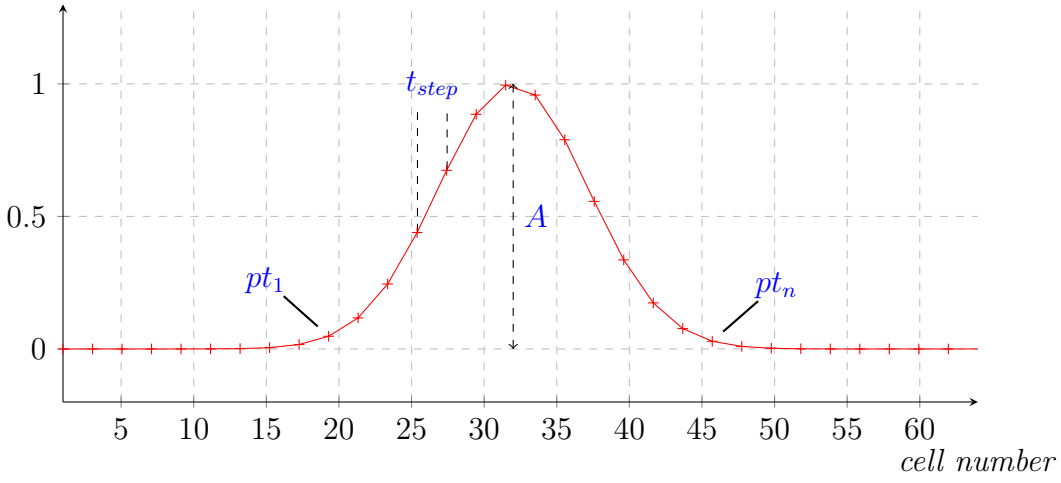


Figure 6.17 – Illustration of the 3 parameters that will be varied during the analysis. The amplitude A of the signal - a lower amplitude should impact the pulse affected by sampling noise ϵ_i only. The number of non-zero points n - from pt_1 to pt_n - without modifying the time step. The sampling step t_{step} without modifying the number of non-zero points.

In order to separate the noise contribution from n_j and n_ϵ in time delay estimation (Equation 6.51), two types of signals are generated:

Signal one s_1 A Gaussian pulse, similar to the pulse shown in Figure 6.15, is impacted by the **sampling time jitter δ_i** only. Using cross-correlation, a time t_1 is extracted. We expect the inverse square root of the variance $1/\sqrt{\text{Var}(t_1)}$ to have the following dependence :

$$\frac{1}{\sqrt{\text{Var}(t_1)}} \propto \frac{\sqrt{n}}{\sigma_j}. \tag{6.52}$$

Signal two s_2 A Gaussian pulse, similar to the pulse shown shown in Figure 6.16, is impacted by the **sampling noise ϵ_i** only. Using cross-correlation, a time t_2 is extracted. We expect the inverse square root of the variance $1/\sqrt{\text{Var}(t_2)}$ to have the following dependence:

$$\frac{1}{\sqrt{\text{Var}(t_2)}} \propto \frac{2A}{\sqrt{n}t_{step}\sigma_n} \tag{6.53}$$

6.6.1 Amplitude dependence

Here we verify the effect of the amplitude dependence on the time extracted from signals affected by:

- **Sampling time jitter δ_i** only - signal 1 —+— in Figure 6.18.
- **Sampling noise ϵ_i** only - signal 2 —+— in Figure 6.18.

In the simulation, we vary the amplitude A, varying from 1 to 100 (in arbitrary units). Additionally, we set:

- The sampling time step t_{step} is fixed.
- The RMS noise σ_n^2 is fixed.
- The RMS of the sampling time jitter σ_j^2 is fixed.
- The Full Width Half Maximum of the signal is fixed and equal to $10 \times t_{step}$ - similar to fixing the number of non-zero points n.

The times t_1 and t_2 are calculated using the cross-correlation for a given template t. The square root of the inverse variance is plotted for both signals in Figure 6.18.

We verify that for *signal 2* (—+—) the square root of the inverse variance $\frac{1}{\sqrt{\text{Var}(t_2)}}$ has a clear linear dependence upon the amplitude A. No dependence is observed for *signal 1* (—+—).

Conclusion Signals with high amplitudes allow for better time resolutions, reducing the error contribution from the sampling noise. This is simply explained by the fact that increasing A improves the signal to noise ratio. This result justifies the search for high gain photodetectors or the use of low noise amplifiers to increase the signal amplitude prior to sampling.

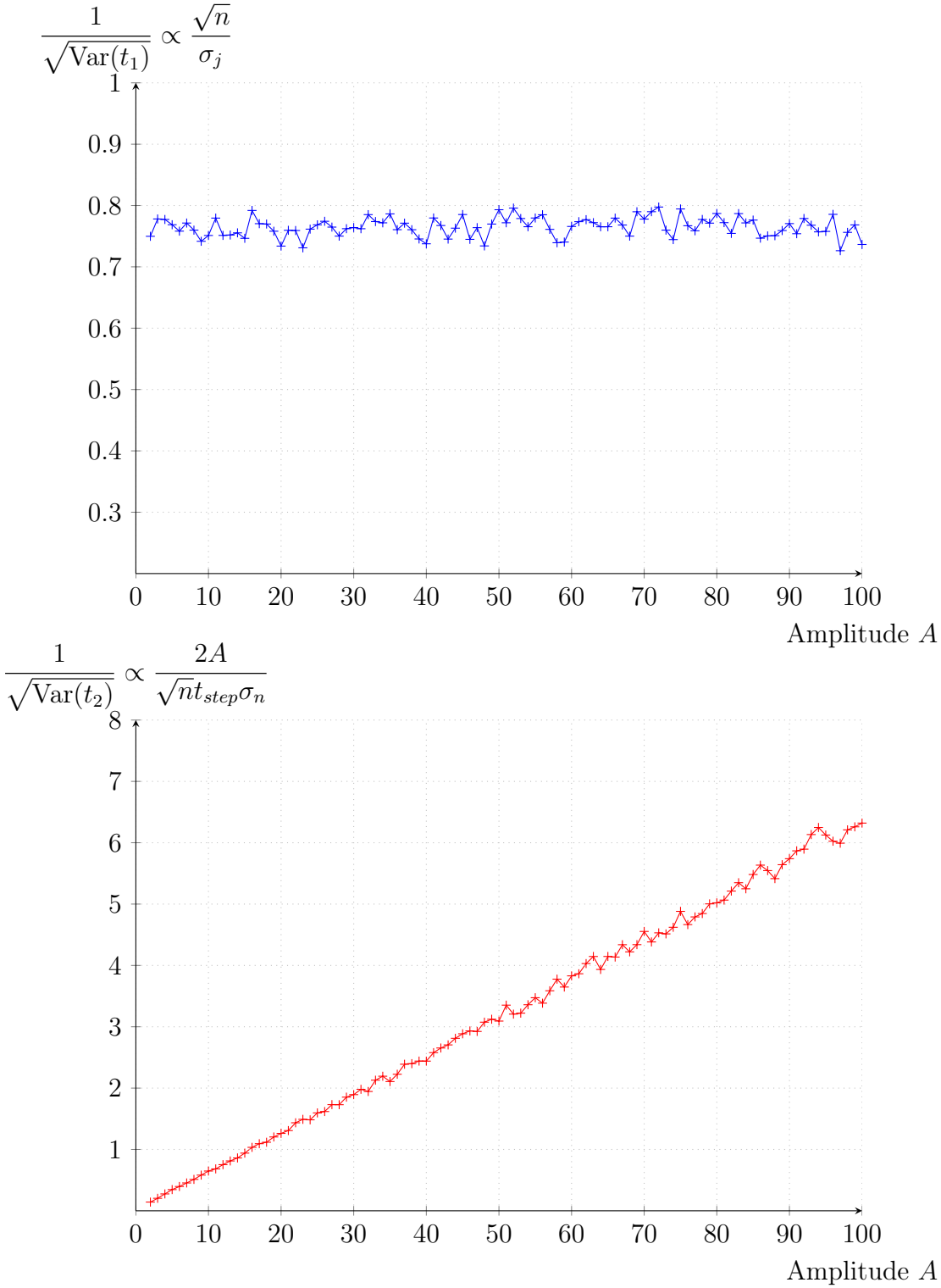


Figure 6.18 – Simulation result of the effect of increasing the amplitude A on the variance of the time delay estimate for Gaussian pulses. As expected no effect is observed on the variance of pulses affected by sampling time jitter (—+—), while a linear dependence is verified for the variance of pulses affected by signal noise (—+—).

6.6.2 Sampling time step dependence

Here we verify the effect of the sampling time step t_{step} on the time extracted from signals affected by:

- **Sampling time jitter δ_i** only - signal 1 —+— in Figure 6.18.
- **Sampling noise ϵ_i** only - signal 2 —+— in Figure 6.18.

In the simulation, we sweep the time step t_{step}, varying from 1 to 100 (in arbitrary units). Additionally, we set:

- The amplitude A is fixed for both signals.
- The RMS noise σ_n^2 is fixed.
- The RMS of the sampling time jitter σ_j^2 is fixed.
- The Full Width Half Maximum of the signal is set equal to $10 \times t_{step}$ - in order to fix the number of non-zero points n [3].

The times t_1 and t_2 are calculated using the cross-correlation to a given template t. The square root of the variance is plotted for both signals in Figure 6.18.

We verify that for *signal 2* (—+—) the square root of the variance $\sqrt{\text{Var}(t_2)}$ has a clear linear dependence upon the time step size t_{step}, while no dependence is observed for *signal 1* (—+—).

Conclusion Here the simulation shows that sampling 10 points every 100ps on a 1ns FWHM signal will give a better precision that for a 10ns FWHM sampled every 1ns. This motivates us in two directions: increase the bandwidth of the signal being sampled together with increasing the sampling speed.

[3] The full width half maximum of the signal is increasing with the time step size, this means that the signal is getting slower.

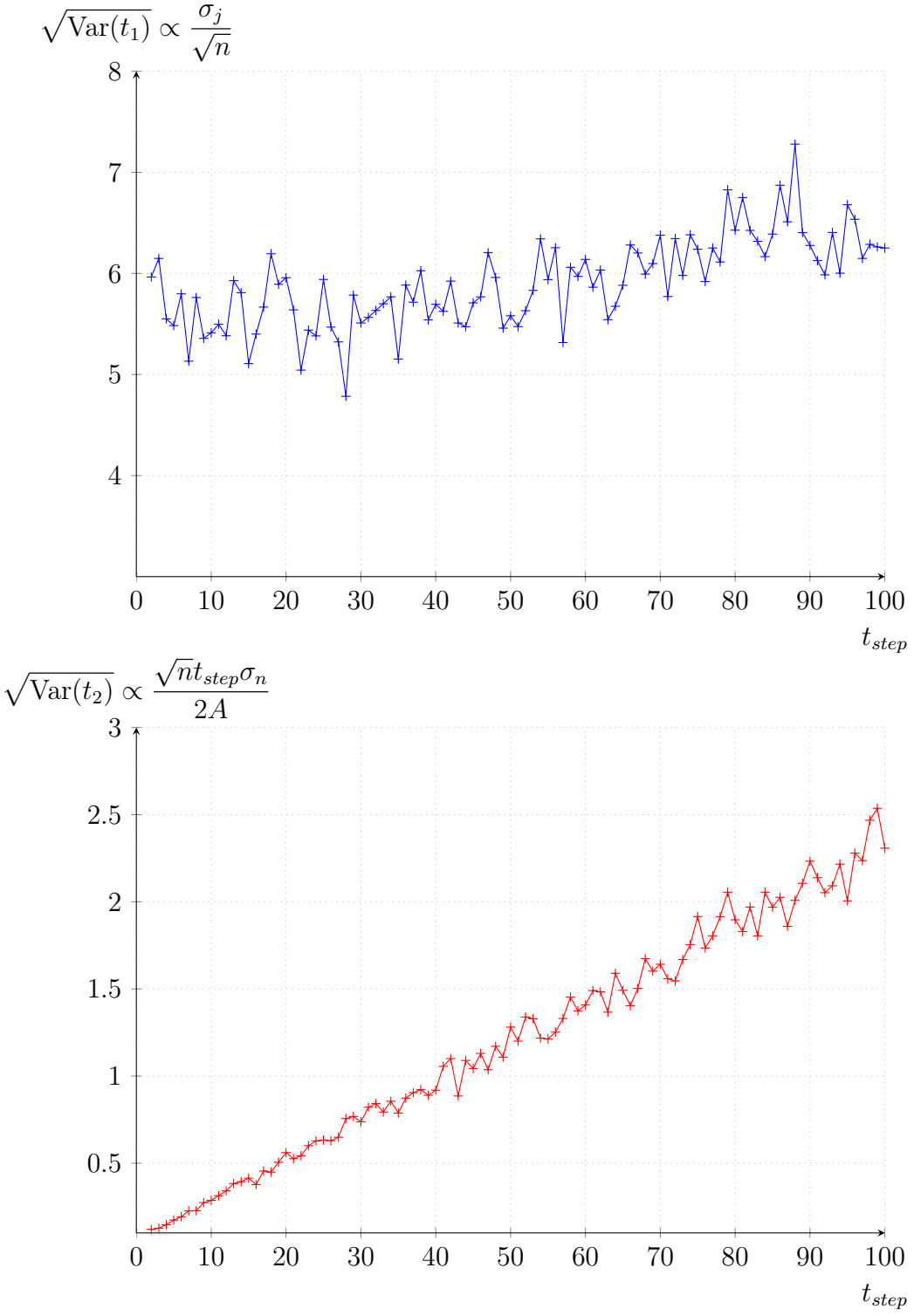


Figure 6.19 – Simulation of the effect of an increase in the sampling time step t_{step} on the variance of the time delay estimate for Gaussian pulses. As expected no effect is observed on the variance of pulses affected by sampling time jitter (—+—) while the linear dependence is verified on the variance of pulses affected by signal noise (—+—).

6.6.3 Number of points and bandwidth dependence

Here we verify the effect of the number of non-zero point n on the time extracted from signals affected by:

- **Sampling time jitter $\boldsymbol{\delta_i}$** only - signal 1 —|— in Figure 6.18.
- **Sampling noise $\boldsymbol{\epsilon_i}$** only - signal 2 —|— in Figure 6.18.

In the simulation, we sweep the increase the number of points n on the signal by increasing the Full Width Half Maximum of the gaussian pulses. The rest of the parameters are set:

- The amplitude A is fixed for both signals.
- The sampling time step t_{step} is fixed for both signals.
- The RMS noise σ_n^2 is fixed.
- The RMS of the sampling time jitter σ_j^2 is fixed.

The times t_1 and t_2 are calculated using the cross-correlation to a given template t. The square root of the variance is plotted for both signals in Figure 6.18.

We verify the expected $1\sqrt{n}$ dependence on pulses affected by sampling time jitter (—|—) and the $\sqrt{n}$ dependence on pulses affected by sampling noise $\mathrm{Var}(e_n)$ (—|—).

Conclusion Experimentally, increasing the number of points on the signal is as simple as increasing the sampling speed. We can observe here, that beyond a point the speed increase is no longer helpful to get a precise measurement as we start integrating noise only.

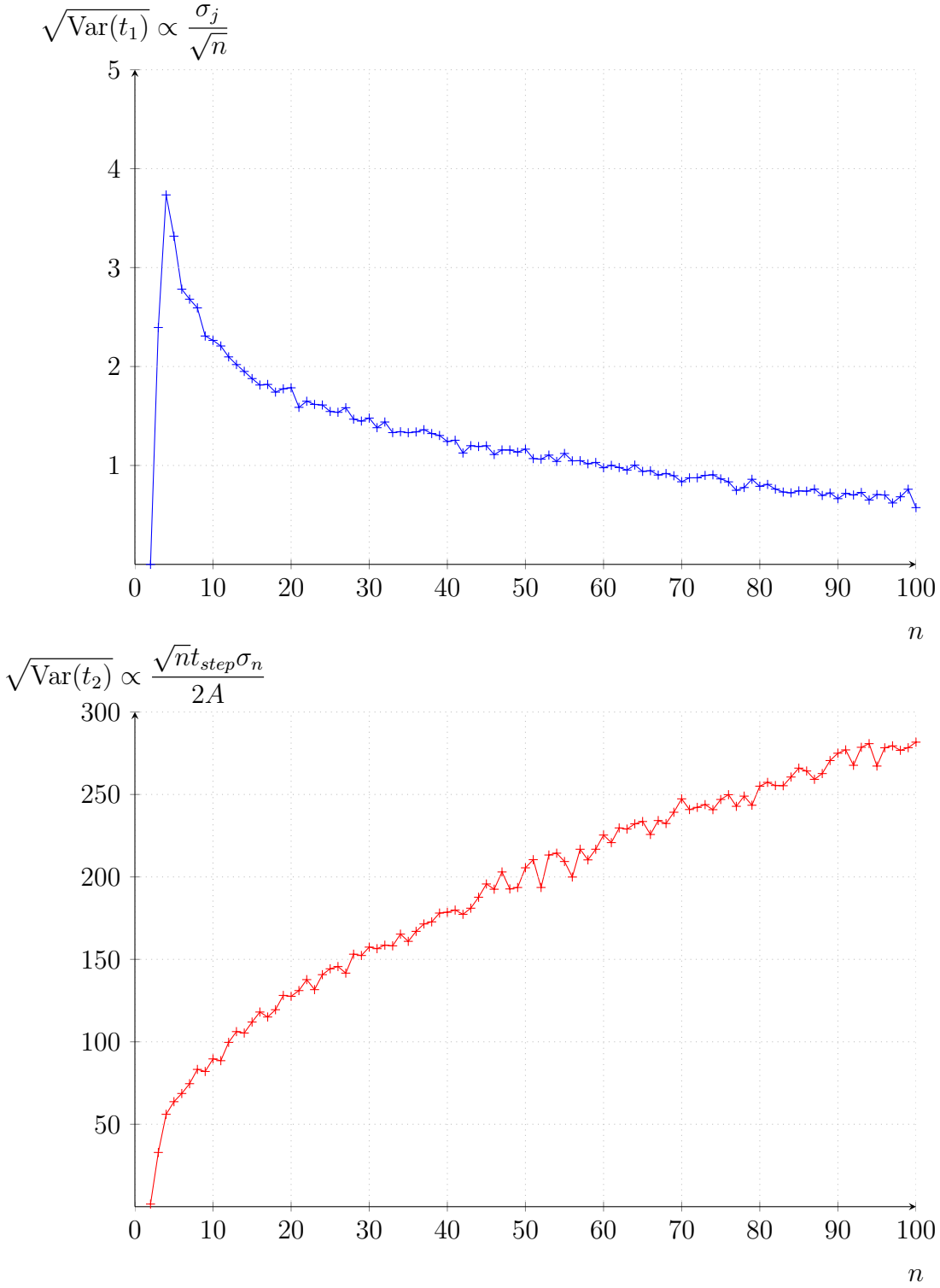


Figure 6.20 – Simulation result of an increase in the number of points n on the variance of the time delay estimate for Gaussian pulses. We verify the expected $1\sqrt{n}$ dependence on pulses affected by sampling time jitter (—+—) and the $\sqrt{n}$ dependence on pulses affected by sampling noise $\mathrm{Var}(e_n)$ (—+—).

6.7 Least mean square algorithm

6.7.1 Introduction

The least mean square algorithm corresponds to calculating the sum of the square differences between the template t and the signal s:

$$l[k] = \sum_{0}^{N-1} (s[i] - t[i+k])^2 \;\; k \in [-N+1..N-1] \tag{6.54}$$

This sum takes its minimum value when $\forall i: \; s[i] \approx t[i-k]$, corresponding to k:

$$k = \frac{t_{lag}}{t_{step}} \tag{6.55}$$

6.7.1.1 Equivalence to cross-correlation

If we write the least-mean square algorithm as finding k such as we minimize the mean square sum we have:

$$\begin{aligned} t_{lag} &= \min_{k \in [-N+1, N-1]} \sum_{0}^{N-1} (s[i] - t[i+k])^2 \\ &= \min_{k \in [-N+1, N-1]} \sum_{0}^{N-1} \left(s[i]^2 - 2s[i]t[i+k] + t[i+k]^2 \right) \\ &= \min_{k \in [-N+1, N-1]} \sum_{0}^{N-1} \left(s[i]^2 + t[i+k]^2 \right) - \sum_{0}^{N-1} 2s[i]t[i+k] \\ &= \max_{k \in [-N+1, N-1]} \sum_{0}^{N-1} 2s[i]t[i+k] \end{aligned} \tag{6.56}$$

Indeed the sum over k of $s[i]^2$ and of $t[i+k]^2$ does not depend on k and does not play a role in the algorithm. So effectively the least-mean square algorithm is equivalent to the cross-correlation [6] and will give similar results.

From a performance point of view, one would prefer the cross-correlation algorithm, yielding to less calculation than the mean square ($2 \times N$ instead of $3 \times N$ for the least-mean square).

6.8 Data interpolation

Most of the other complex algorithms that extract the time of a signal: the constant fraction method for example, often relies on an interpolation of the data to improve its precision.

Interpolation relies on the minimization of point of a template $t[j]$ to the signal samples $s[i]$. The minimization is done by:

$$t_{lag} = \min_{\forall j} \sum_{0}^{N-1} \left(s[i] - t[j]\right)^2 \tag{6.57}$$

And when the template is taken similar to the signal waveform s, the data interpolation is nothing else than the **least mean square algorithm** and thus will give similar results to the cross-correlation.

6.9 Conclusion

To summarize, we have obtained here a clear sense of the limiting factor for extracting precise time on a sampled signal. They are expressed simply via the standard deviation σ_t, or the variance given by the cross-correlation algorithm:

$$\boldsymbol{\sigma_t \sim \sqrt{\frac{n t_{step}^2 \sigma_n^2}{4A^2} + \frac{\sigma_j^2}{n}}} \tag{6.58}$$

This result is in good agreement with the results reported in [3]; precise timing relies on:

- High signal to noise ratio.
- Lower sampling time jitter.
- High analog bandwidth.
- Fast signals.

These result motivates our fast, high analog bandwidth and low noise sampling ASICs approach to readout fast photodetectors.

Bibliography

[1] MATLAB. *version 7.10.0 (R2010a)*. The MathWorks Inc., Natick, Massachusetts, 2010.

[2] BG Taylor. *TTC distribution for LHC detectors*. *Nuclear Science, IEEE Transactions on*, 45(3):821–828, 1998.

[3] Jean-Francois Genat, Gary Varner, Fukun Tang, and Henry Frisch. *Signal processing for picosecond resolution timing measurements*. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 607(2):387–393, 2009.

[4] G Clifford Carter. *Coherence and time delay estimation*. *Proceedings of the IEEE*, 75(2):236–255, 1987.

[5] Charles Knapp and Glifford Carter. *The generalized correlation method for estimation of time delay*. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 24(4):320–327, 1976.

[6] Svante Björklund. *A survey and comparison of time-delay estimation methods in linear systems*. Univ., 2003.

Part III

Picosecond readout ASICs design and results

7 ASICs

Contents

7.1 Introduction

In Chapter 8 we present the SamPic chip, which was designed, fabricated and tested in Saclay. The purpose of this chip is to measure the time utilizing the techniques for time delay estimation presented in Chapter 6. The goal here is therefore to design a device capable of very fast sampling, with respect to precise reference time, and recording both samples and the time reference. The concept was introduced in Chapter 6, but is shown again here in Figure 7.1.

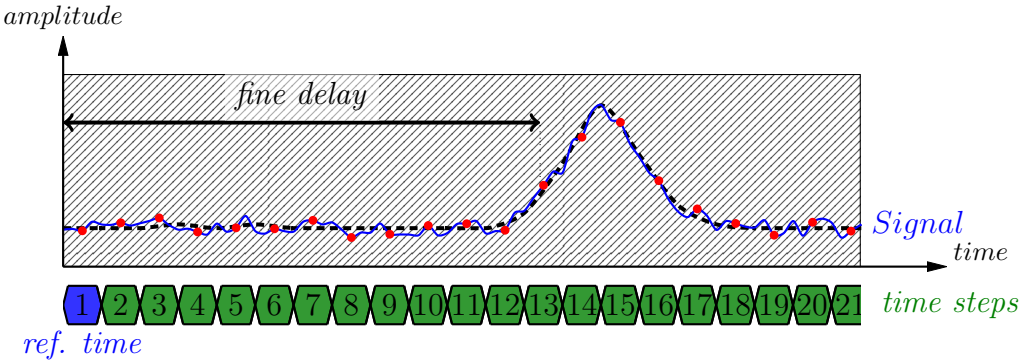


Figure 7.1 – The objective of the ASIC is to sample (red dots) the continuous signal (in blue) at precise time steps (in green). The sampling will be done with respect to a perfectly known clock, providing excellent time resolution with respect to the first sample point *ref. time* (in blue). Both samples and the reference time are recorded.

The design is done with special care to minimize the sampling noise and the sampling time jitter, which are the key parameters for precise time measurement (in Chapter 6).

7.2 Psec ASIC family

The SamPic chip design was started during the second year of this thesis. It takes into account lessons learnt during my contribution to the PSEC chips family when I was with the University of Chicago. The PSEC-4 chip, the last of the family achieved very good performances, both in sampling speed and noise. It will not be presented here, though its characteristics can be found in the references:[1], [2], [3],[4].

The SamPic design is not a simple evolution from the Psec4 chip, which was designed using the IBM CMOS $0.13\mu m$. Different design option choices have been taken to make it more robust and to get better performances. One of the major differences was also the choice of a microelectronic technology, with a larger gate width, offering lower leakage, higher dynamic range at a price four times smaller.

Moreover, the SamPic chip has been designed as a complete timing system, already useable in experiments with high rates.

7.3 Conclusion

The SamPic and Psec chips were developed for precise timing, but specially in order to record signals coming from the LAPPD photo-detector (Chapter 5) or the timing detector of the AFP station (Chapter 3). Therefore, a special attention was paid to the detector characteristics during all steps of the design, in order to develop electronics compatible and integrable directly into experiments.

Bibliography

[1] H. Grabas, E. Oberla, K. Attenkoffer, M. Bogdan, H.J. Frisch, J. F Genat, R. Northrop, E.N. May, G.S. Varner, and M. Wetstein. *Development of large area, pico-second resolution photo-detectors and associated readout electronics*. In *Advancements in Nuclear Instrumentation Measurement Methods and their Applications (ANIMMA), 2011 2nd International Conference on*, pages 3–5, 2011.

[2] M. Bogdan, H.J. Frisch, J.-F.C. Genat, H. Grabas, M.K. Heintz, S. Meehan, E. Oberla, L.L. Ruckman, Fukun Tang, and G.S. Varner. *Development of a 20 GS/s sampler chip in 130nm CMOS technology*. In *Nuclear Science Symposium Conference Record (NSS/MIC), 2009 IEEE*, pages 1929–1931, 2009.

[3] Michael Cooney, Matt Andrew, Kurtis Nishimura, Larry Ruckman, Gary Varner, Hervé Grabas, Eric Oberla, Jean-Francois Genat, and Large Area Picosecond Photodetector Collaboration. *Multipurpose Test Structures and Process Characterization using 0.13 µm CMOS: The CHAMP ASIC*. *Physics Procedia*, 37:1699–1706, 2012.

[4] E Oberla, H Grabas, M Bogdan, H Frisch, JF Genat, K Nishimura, G Varner, and A Wong. *A 4-Channel Waveform Sampling ASIC in 0.13 µm CMOS for front-end Readout of Large-Area Micro-Channel Plate Detectors*. *Physics Procedia*, 37:1690–1698, 2012.

8 SamPic ASIC

Contents

8.1 Introduction

As described in Chapter 7, we present here the work on the SamPic ASIC design. SamPic stands for Sampler for Picosecond time pick-off. As the name implies, the objective of this ASIC is to measure time precisely, thus to sample ultra-fast signals with low noise and jitter, referenced to a precise absolute time [1].

8.2 What we are trying to achieve

The SamPic chip detects, timestamps, samples and converts fast analog pulses. The principle of its operation is illustrated in Figure 10.2.

The SamPic chip can process 16 different input signals in as many channels, (Section 8.4). The general architecture of the chip can be seen in Figure 8.3.

Time references Inside the chip, the sampling steps are defined by a *delay line*, locked to an external reference clock Ck_{ref} (Section 8.7).

Two time base are used to keep track of the elapsed time:

- A coarse *Gray counter* sampling the external, precise, reference clock Ck_{ref}.
- A fine time base with the granularity of the delay line.

Channels In each channel, the threshold crossing is detected by the internal *discriminator* (Section 8.10).

Each channel continuously stores analog samples of the input signal in a FIFO buffer of *analog sampling cells* (Section 8.6).

When a trigger happens the writing cycles in the *analog sampling cells* are stopped and a *hit* is recorded (Section 8.10).

- The *DLL position* and the *coarse counter* values are latched.
- The *analog sampling cells* values are digitized by the internal *ADC* (Section 8.9).

Triggering, conversion and readout of the cells for each channel is managed by the *Channel controller*.

Chip readout The *Chip controller* provides trigger, conversion and readout signal to each channel (Section 8.12).

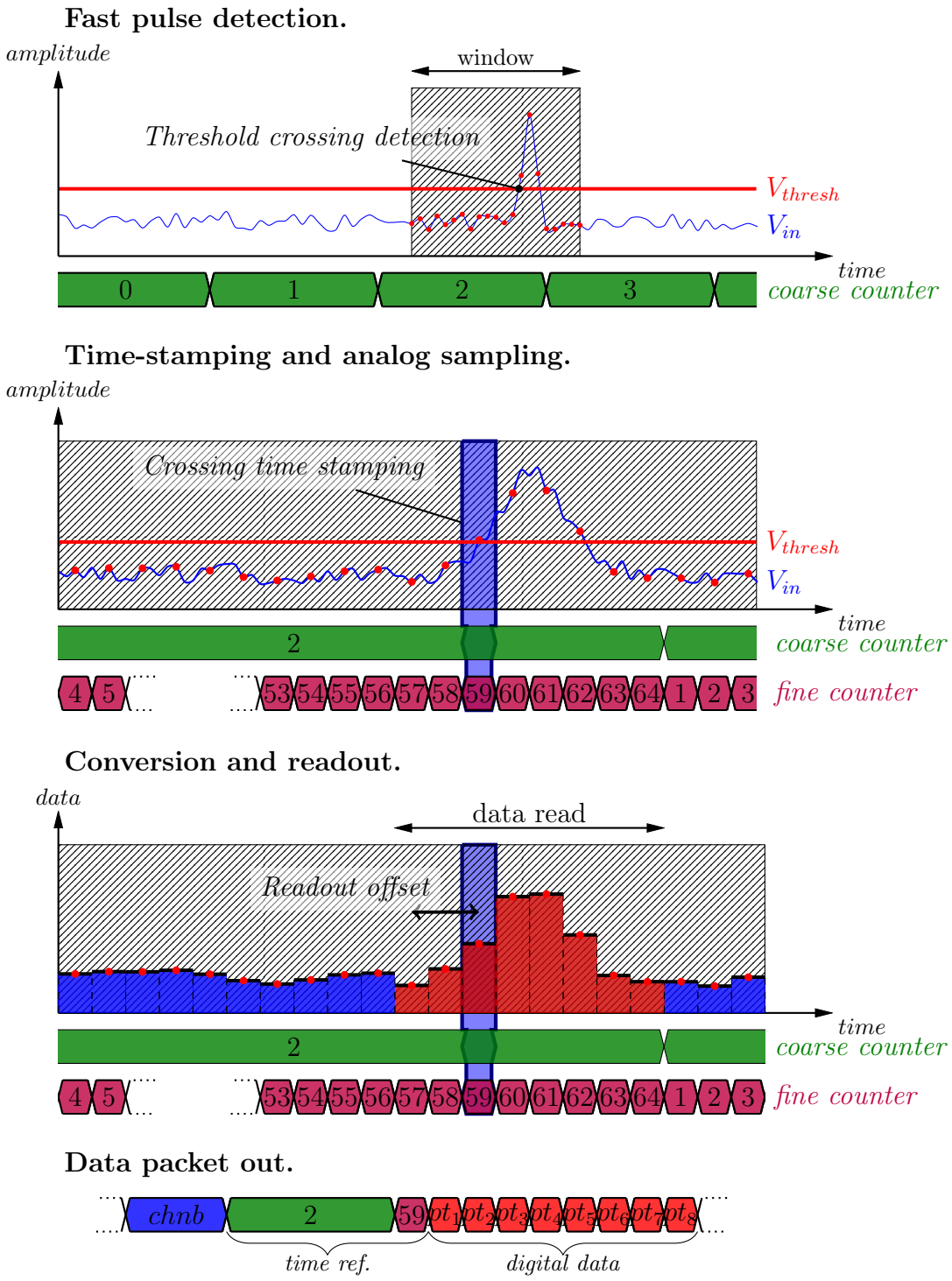


Figure 8.1 – SamPic's principle is illustrated here: fast pulse detection, time-stamping and analog sampling, conversion and readout.

8.3 Technology used

SamPic is fabricated using the 0.18μm High Voltage CMOS design kit provided by Austria Micro System [2] - however, the high voltage part of the kit is not used. The details of the characteristics and performances of this technology are company confidential. We will only illustrate the transistor structure and metal layers available, shown in Figure 8.2. In this technology, the top level metal, being 10 times thicker than the lower ones is much less resistive and is used for routing of the input signals.

8.3.1 The 0.18μm CMOS process

The C18 Mixed-Signal process is the base process for AMS advanced 0.18 μm process family. C18 is equivalent to IBM's industry proven foundry process technology CMOS7RF and offers highest integration density up to 118kGates/mm^2 at up to 7 levels of metal, supply voltages from 1.8 V to 5.0 V and ESD protection cells with up to 8 kV HBM level. The C18 process is manufactured in a state of the art 200 mm fabrication process ensuring very low defect densities and high yields [2].

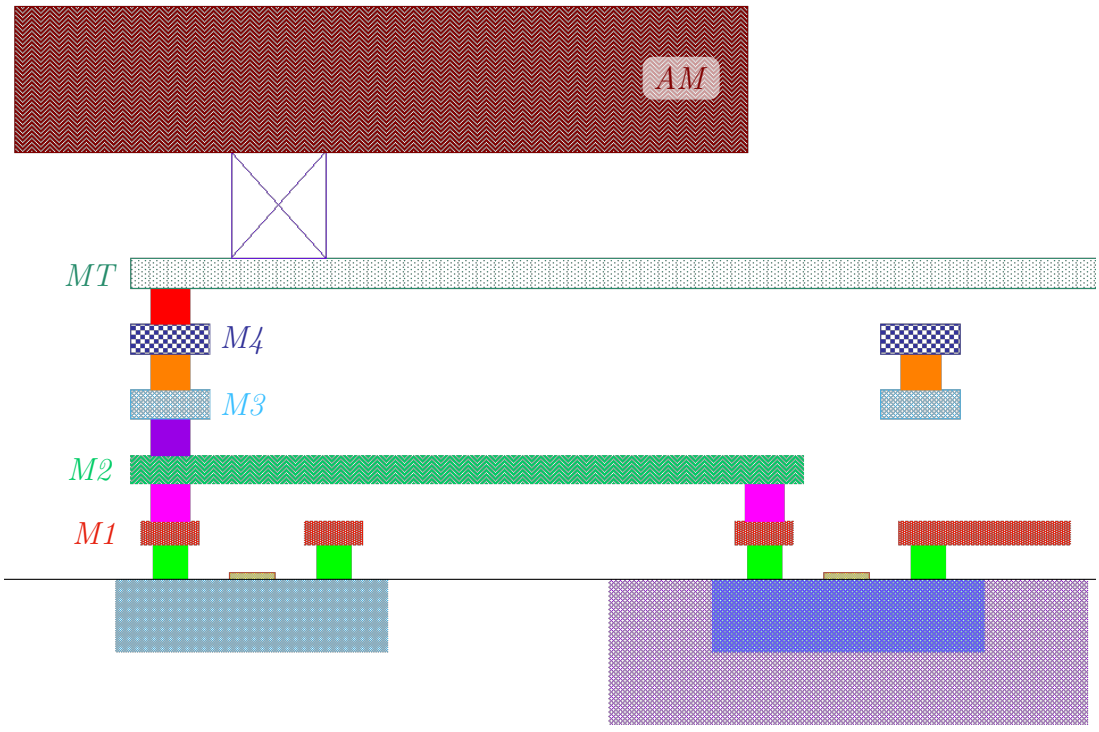


Figure 8.2 – Illustration of the metal layer structure in the AMS 0.18 μm CMOS technology used for SamPic. 6 layers of metals are available, with one thick, low resistivity top layer. Illustration is not to scale.

The industry leading design environment (hitkit) includes silicon-qualified digital, analog and RF library elements and a complete set of low voltage devices (1.8 V and 5.0 V). Fully characterized simulation models including noise, matching and

parasitic elements, extraction and verification run sets as well as automatic layout device generators complete the C18 hitkit offering. Hence product developers are enabled with a plug-and-play tool set which facilitates *first time right* designs [2].

The use of the 0.18 μm High Voltage CMOS design kit provided by Austria Micro System, is therefore motivated by a compromise of the following criterions:

- **Cost** About 1000 $ per mm^2.
- **Density** Up to 118 kGates/mm^2.
- **Design kit** Excellent AMS standards.
- **Leakage** Smaller compared to the 0.13 μm option.

8.4 General architecture

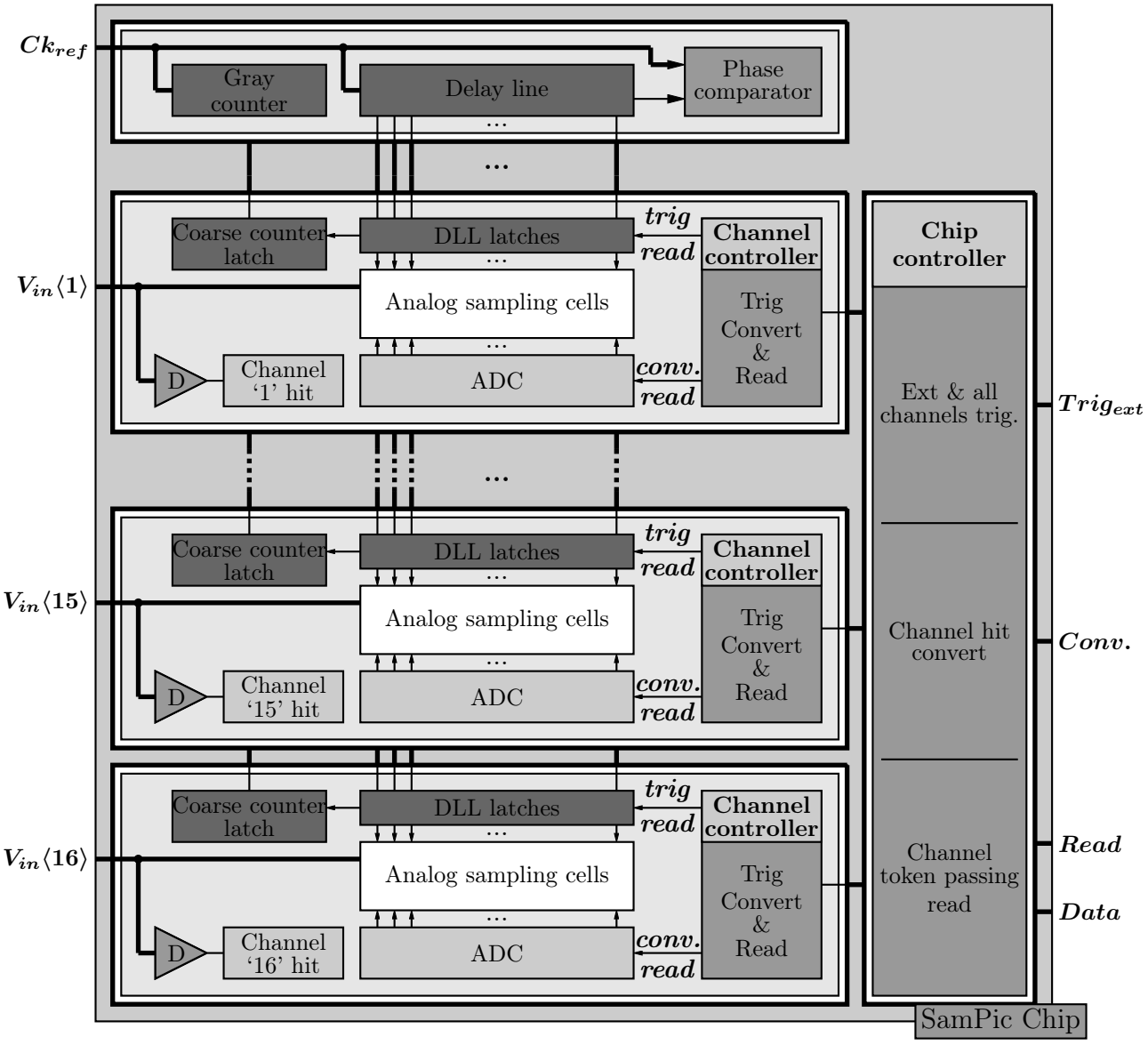


Figure 8.3 – Top level block diagram of the SamPic chip.

8.4.1 Discriminator

There is one discriminator per channel. The role of the discriminator is to detect a fast pulse at the channel *input*. It permits to define a window around the signal-of-interest (threshold crossing) where 64 samples of the signal will be logged as shown in Figure 8.4. The details are described in Section 8.10.

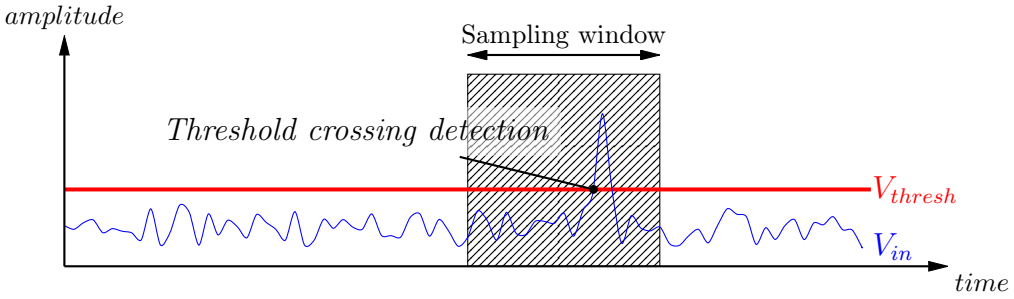


Figure 8.4 – The role of the discriminator inside each of SamPic's channels is to detect fast input pulses.

8.4.2 Sampling cell array

The analog buffering inside each SamPic channel is performed with an array of 64 sampling cells. The role of these cells is to sequentially track-and-hold the input signal of the channel and continuously store the 64 previous data samples as shown in Figure 8.5. The details are described in Section 8.6.

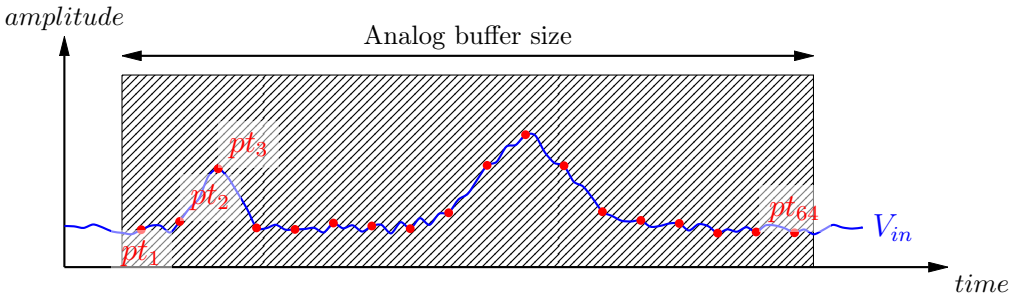


Figure 8.5 – The role of the SamPic sampling cells is to continuously buffer analog samples of the input signal.

8.4.3 Slow and Fast time base

The role of the slow and fast clock inside SamPic is to provide a time reference for the events sampled inside each channel, as shown in Figure 8.6. The fast time base provides a reference for each point taken in SamPic while the slow time base, provided by the slow clock of SamPic, provides a reference for each "frame" (group of 64 samples) taken in SamPic.

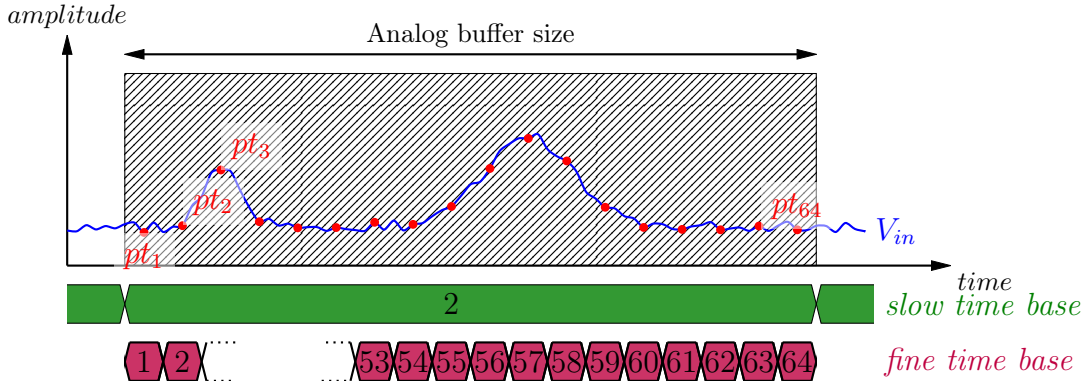


Figure 8.6 – The slow and fast time base in SamPic provide accurate time reference for all sampled points.

8.4.4 Delay locked loop

A 64-bit delay locked loop (DLL), which total delay is servo-controlled to the slow time base duration is used to provide the 64 signals used to control the sampling in the SamPic. The same signals are used to determine the fast time base. The delay locked loop guarantees that the sampling time steps are accurate and equally spaced, as shown in Figure 8.7. The details are described in Section 8.7.

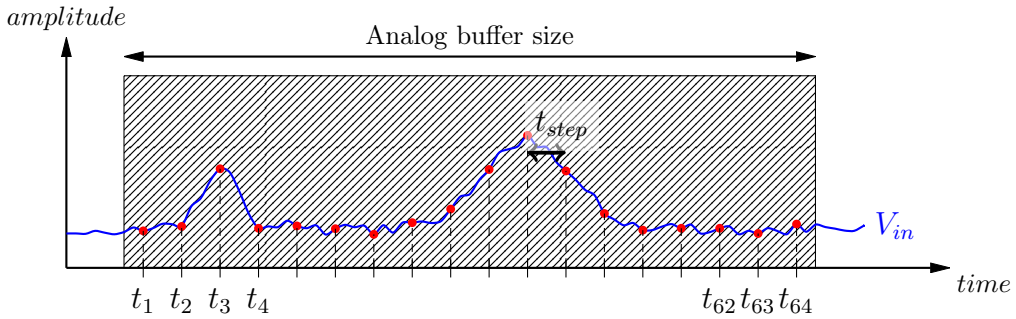


Figure 8.7 – The role of the DLL of SamPic is to provides for fast accurate sampling times.

8.4.5 Analog to digital converter

Once sampled, the analog values will be converted with a resolution of 11-bit. This is the role of the analog-to-digital (ADC) converter present in each channel, as shown in Figure 8.8. The details are described in Section 8.9.

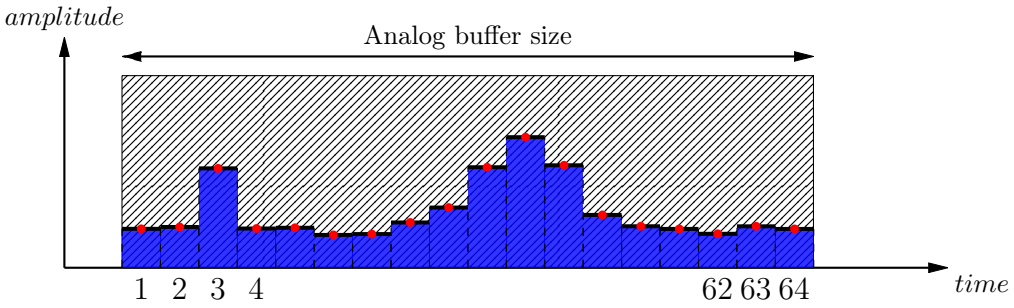


Figure 8.8 – The role of the ADC of each of the SamPic's channels is to convert the analog sampled value to digital values.

8.4.6 Region of interest readout

Once converted, the data of the 16 converted channels are read by the acquisition. In each channel the region-of-interest (ROI) finder allows for smart readout of valuable data only (around the trigger) hence speeding up the readout process, as shown in Figure 8.9.

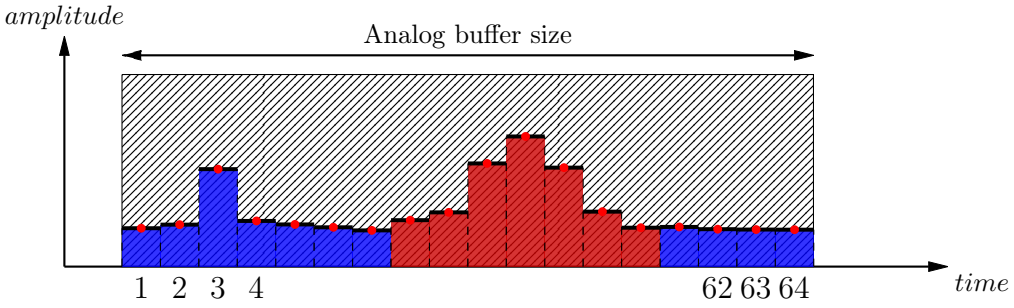


Figure 8.9 – The role of the Region-Of-Interest readout is to read only data of interest and therefore speed-up the readout process.

8.5 Channel description

The sampling channel consists of an array of 64 identical memory cells housing the following: an analog sampling cell (Section 8.6), Wilkinson analog-to-digital converter (Section 8.9), 11-bit digital data storage (Section 8.9) and trigger generator for the sampling cell. The channel block diagram is shown in Figure 8.10. The details are described in Section 8.10.

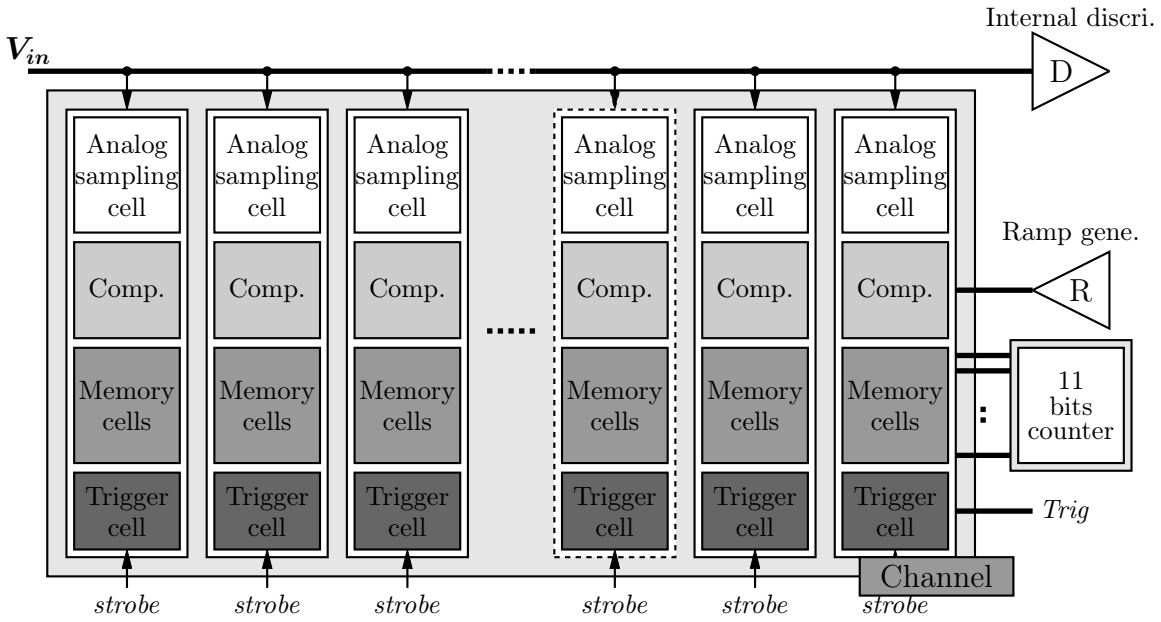


Figure 8.10 – Block diagram of a SamPic channel. Each consists of a group of 64 memory cells, each housing an analog sampling cell, a Wilkinson AtoD converter, a digital storage cell and a trigger cell.

8.6 Sampling cells

The sampling in each channel is performed with an array of 64 sampling cells, which is a rather small value. This decision was made as a conclusion of the Chicago Timing workshop [3]: a smaller number of sampling cell increasing the channel compacity, hence the bandwidth. Each cell is individually controlled by a *trigger cell* (see Section 8.10) using the *strobe* signal from the delay locked loop (see Section 8.7) and the *trigger* signal. The architecture is shown in Figure 8.11.

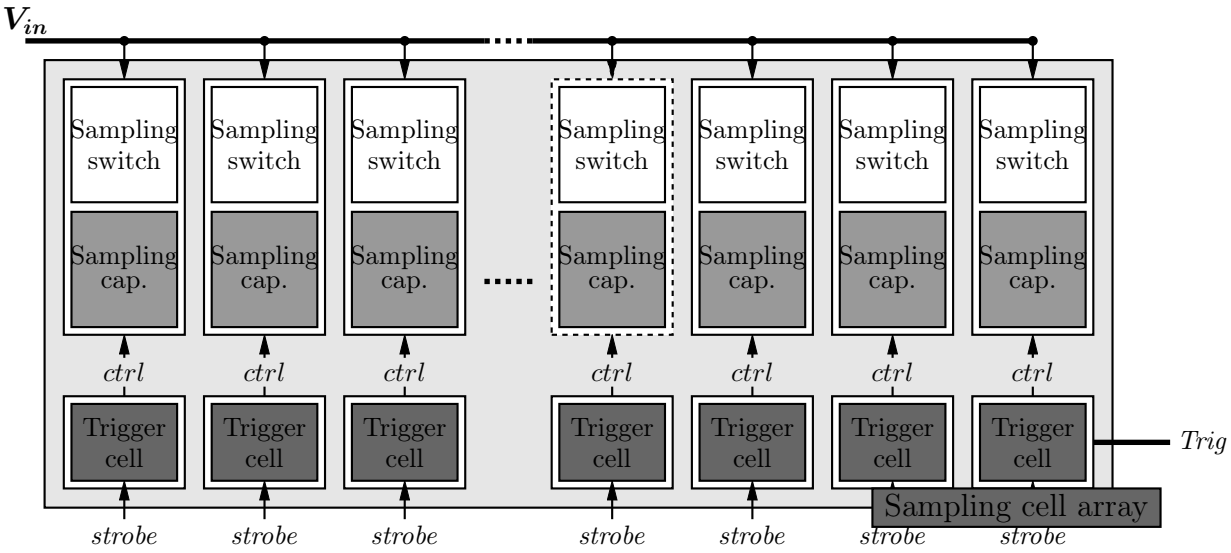


Figure 8.11 – Block diagram of the sampling cells in the SamPic channel. There are 64 sampling cells in every channels. Each sampling cell consists of a sampling switch and a sampling capacitor.

8.6.1 Sampling principle - one sampling cell

The principle of sampling inside *one sampling cell* is described in Figure 8.12. The *control* signal from the *trigger cell*, puts the cell in *tracking* mode when it is high. During the track mode the voltage in the cell follows the input voltage V_{in}. When the *control* voltage goes low, the sampling cell stops tracking the input voltage and hold its value. The principle is illustrated in Figure 8.12.

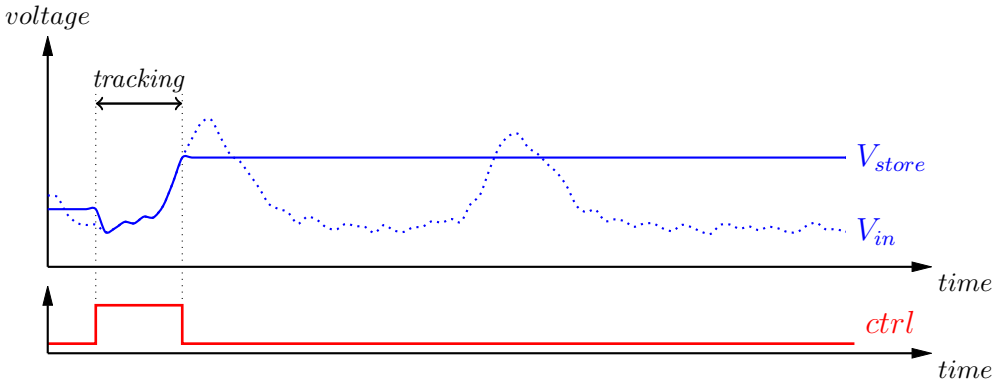


Figure 8.12 – Principle of sampling in *one sampling cell*. During the track mode the voltage in the cell follows the input voltage V_{in}. When the *control* voltage goes low, the sampling cell stops tracking the input voltage and holds its value.

8.6.2 Sampling with an array of cells

The principle of multiple sampling cells connected to the same input line is illustrated in Figure 8.13. Here, four cells are connected and controlled by four identical *control* signals shifted in time by t_{step}. This allows to sample four equally spaced data points (pt_1 to pt_4) on the input signal V_{in}.

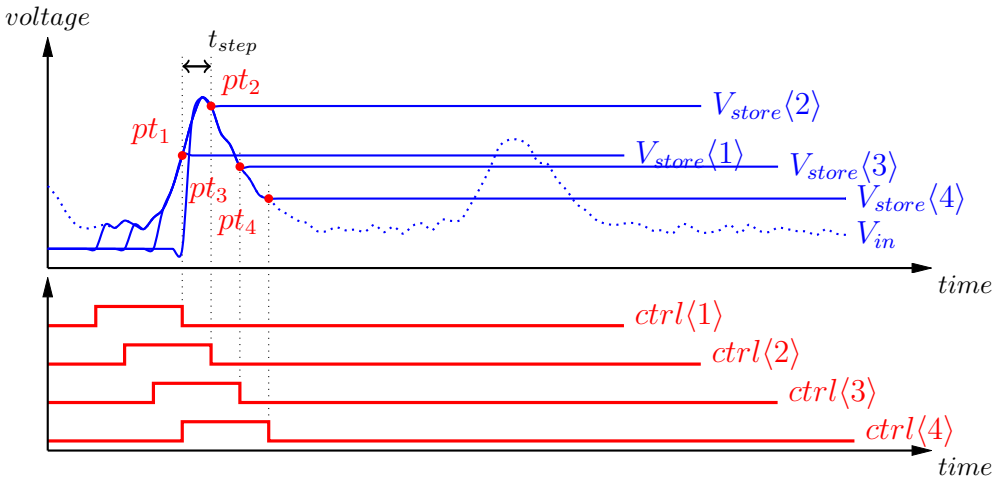


Figure 8.13 – Principle of sampling using a group of 4 cells. Here 4 cells are connected and controlled by four identical *control* signal shifted in time by t_{step}.

8.6.3 Buffering with an array of cells

In order to continuously buffer the analog input signal V_{in}, the sampling cells are written over when all the cells have been written, making the buffer circular. This is illustrated in Figure 8.14, with 6 cells. To ensure continuous buffering in the cell array, the *control* command for the sampling cells are repeated over time. If we have nb_{cell} in the buffer, the repetition time t_{rep} must be:

$$t_{rep} = nb_{cell} \times t_{step} \tag{8.1}$$

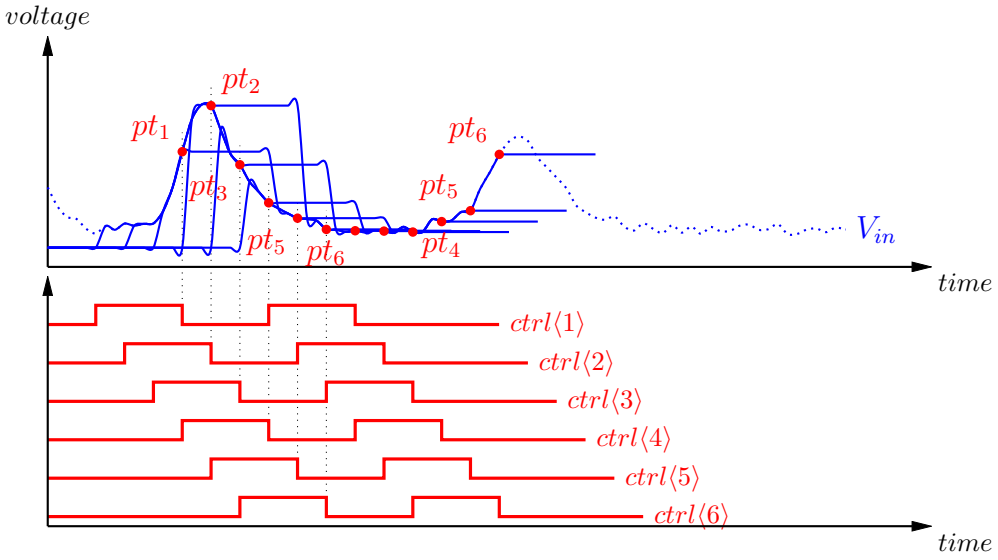


Figure 8.14 – Principle of analog buffering with an array of analog sampling cells. In order to continuously buffer the analog input signal V_{in}, the sampling cells are written over when all the cells are full, making the buffer circular.

8.6.4 Dead buffer cells

Inside the buffer some cells are in tracking mode during $t_{tracking}$, while the others are holding their sampled value. The number of cells in *tracking* mode nb_{track} is given by:

$$nb_{track} = \lceil \frac{t_{tracking}}{t_{step}} \rceil \tag{8.2}$$

The cells in *track* mode inside the buffer do not hold valuable data since they are all connected to the *input* line and share the value $V_{in}(t)$. The rest of the cells $nb_{cell} - nb_{track}$ are in *hold* mode and buffer the previous value of V_{in}. To efficiently buffer the input signal, the number of cells in *track* mode nb_{track} needs to be a small fraction only of the total cell number nb_{cell}.

This principle is illustrated in Figure 8.15. In this example, 5 buffer cells are simultaneously tracking the input signal, while $64 - 5 = 59$ cells are buffering previous input states. If sampling is unconditionally stopped, this 5 buffers cells are simultaneously going from *tracking mode* to *hold mode*, holding the same value, which is what we consider *dead cells*.

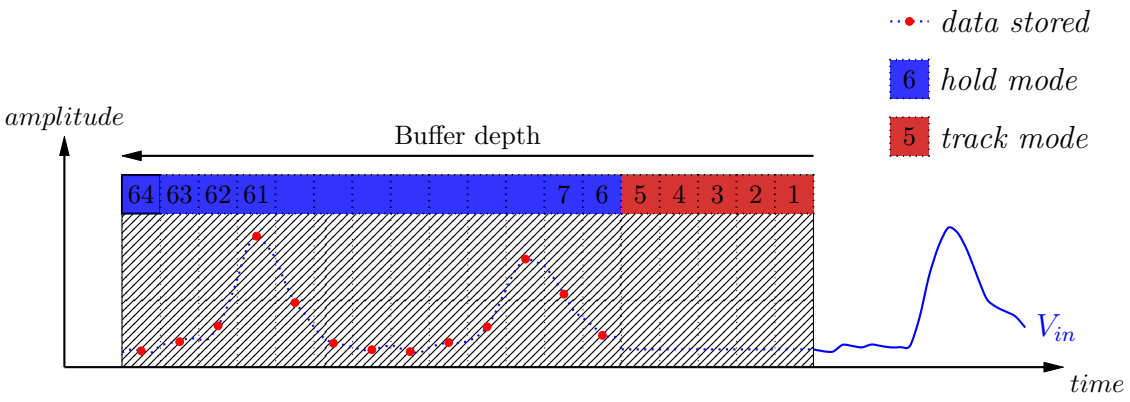


Figure 8.15 – Principle of analog buffering with an array of analog sampling cells.

We can see that the *dead cells* are a limitation to the analog sampling cell buffer, especially for short buffer lenght.

In Section 8.10, a method is presented, which completely removes dead cells. However, the last samples are taken after the trigger signal only.

8.6.5 Sampling sequence

A sampling sequence is presented for illustration in Figure 8.16 for 3 consecutive time steps t_{step}. In this illustration, 5 cells are constantly *tracking* the input signal, ie: $t_{tracking} = 5 * t_{step}$ and appear as *dead cells* in the sampling array.

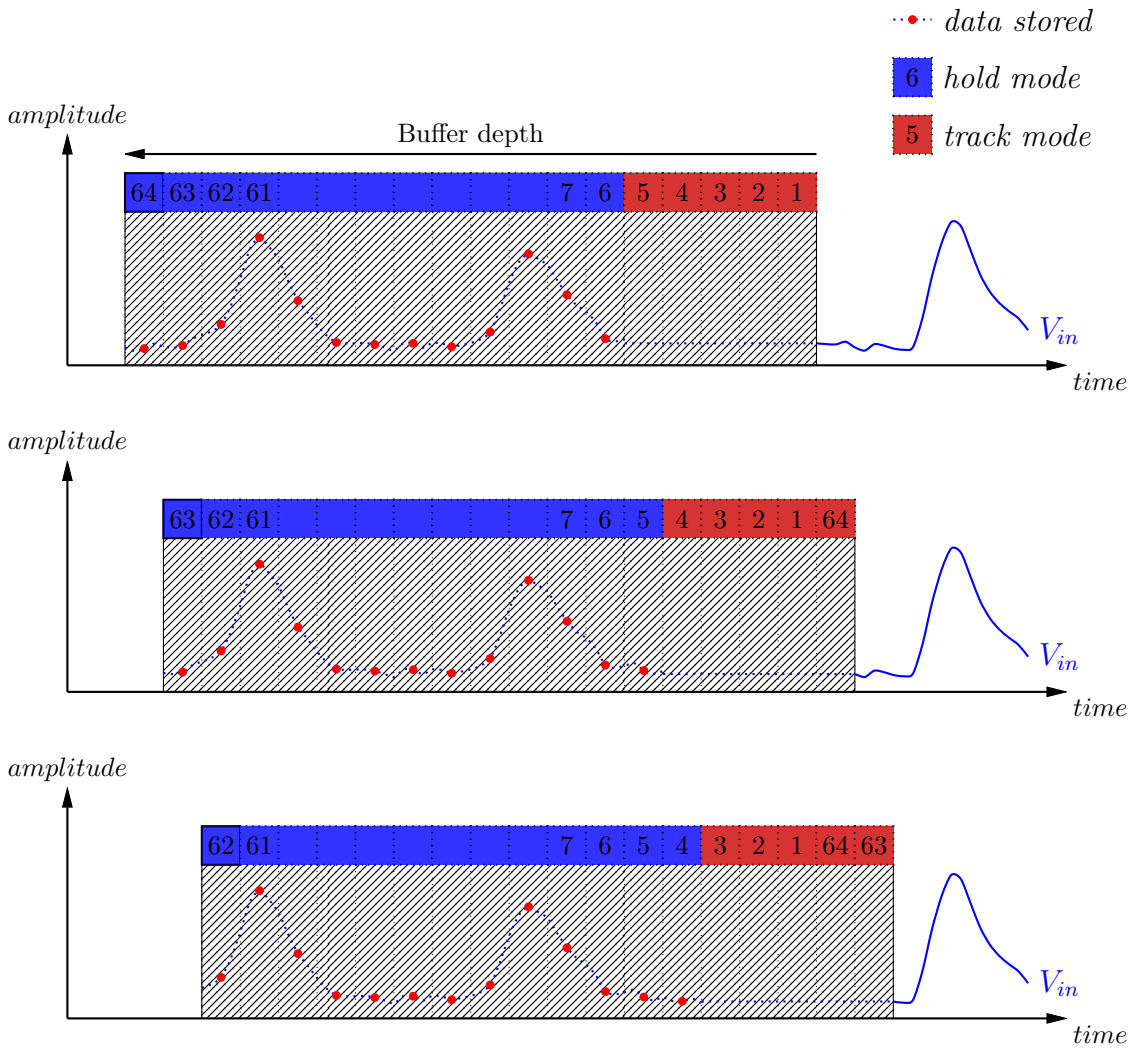


Figure 8.16 – Illustration of the sequence of buffering for 3 time steps t_{step} in an analog buffer of 64 cells with 5 cells connected to the input bus.

8.6.6 SamPic buffer length

Inside SamPic we have chosen an array of 64 analog sampling cells. This number is chosen to be a power of 2, to simplify addressing, and encoding in the chip which uses binary code. A number 64 gives us a sufficiently deep buffer size to record enough data points on a fast signal without being too long and increasing the space, power and readout time requirements for the chip. There are no dead buffer cells in the sampling array (see Section 8.10).

8.6.7 Sampling depth

An illustration of SamPic sampling depth is illustrated in Figure 10.17. Inside each SamPic's channel we can buffer 64 points on the input signal. All the previous data points taken will be overwritten and lost. In time this gives us a buffer depth t_{depth} of:

$$t_{depth} = t_{step} \times 64 \tag{8.3}$$

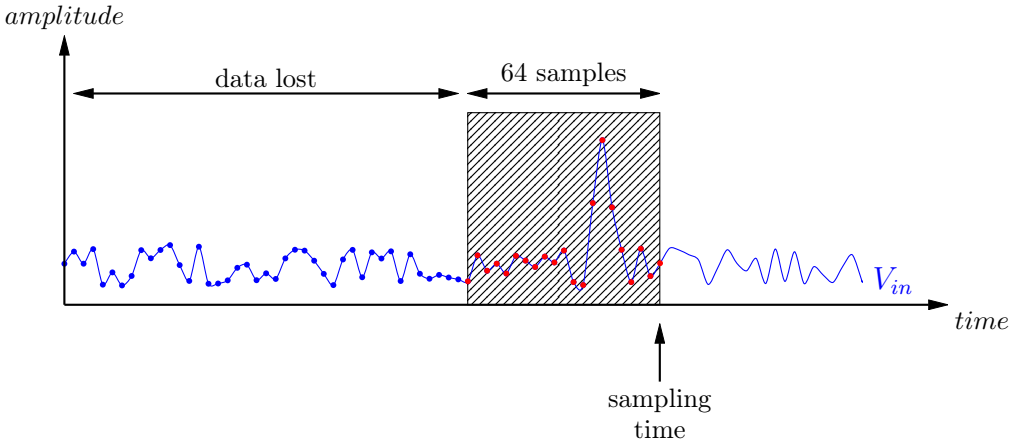


Figure 8.17 – Illustration of SamPic sampling principle and the 64 data point buffer depth. 64 points (red) are being continuously taken on the signal V_{in} after the sampling time. Older, previously taken, points (blue) are overwritten and lost as the time of sampling advances.

8.6.8 Sampling cell design

The simplest sampling cell possible is showed in Figure 8.18a. It is made of a switch and a sampling capacitor. During *tracking*, the switch is closed on the capacitor and open to *hold* the input value. It is important to note that the sampling happens precisely when the switch state changes from closed to open.

(a) Theoretical sampling cell. (b) Sampling cell with transistor switch.

Figure 8.18 – Schematic view of sampling cells, theoretical and as implemented with a CMOS technology.

The switch of Figure 8.18a can be made of an NMOS and a PMOS connected in parallel (see Figure 8.18b).

8.6.9 Sampling noise and quantization error

The noise sampled on the capacitor C_S is given by: $v_n = \sqrt{k_B T / C_S}$ [4] and is plotted in blue (—●—) in Figure 8.19 as a function of the sampling capacitor value. Along with this plot we print the quantization errors for a digitization on 12 (—■—), 11 (—▲—) and 10 bits (—◆—) over a 1.8 V range, which can be calculated as $LSB/\sqrt{12}$.

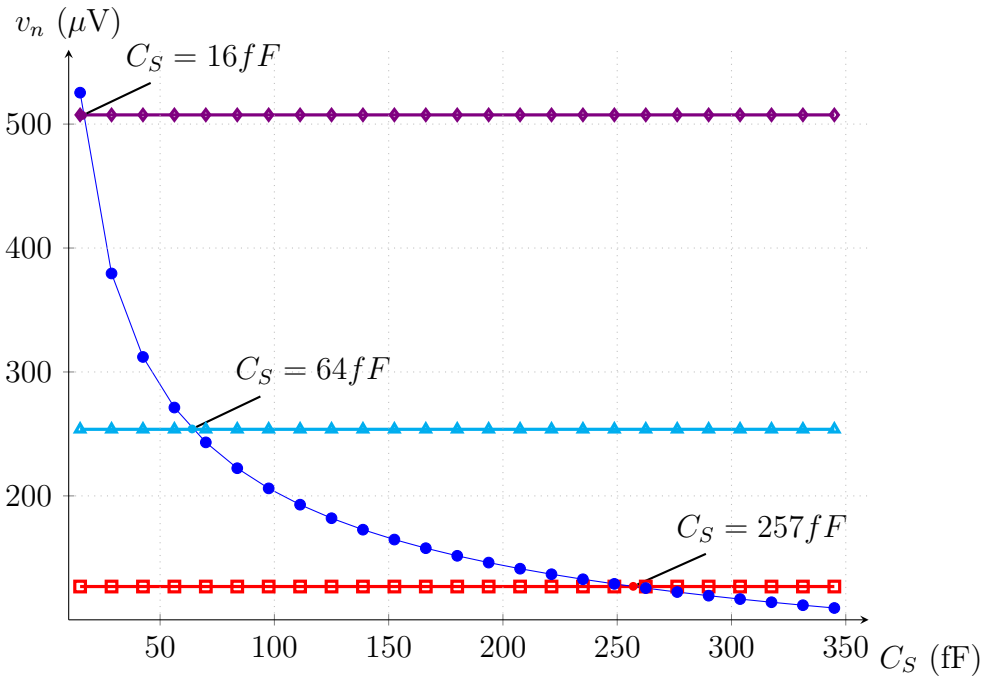


Figure 8.19 – Plot of the sampling noise value $k_B TC$ (—●—) as a function of the sampling capacitor. The quantization noise levels for 12- (—■—), 11- (—▲—) and 10-bits (—◆—) are displayed and their intersection with the noise curve printed.

Ignoring other noise or distortion sources, considering an 11-bit ADC conversion, if $C_S \leq 64$ fF we are dominated by the *sampling noise*. If $C_S \geq 64$ fF, we are dominated by quantization errors.

Table 8.1 – Minimum values of sampling capacitor for sampling noise levels below the 10-, 11- and 12-bits quantization errors on a 1.8 voltage range.

Number of bits	Sampling capacitor value
10	16 fF
11	64 fF
12	257 fF

In conclusion, we have chosen to use a sampling capacitor $C = 35$ fF that limits the verical resolution to 10- to 11-bits. This choice is a trade-off between low noise and high bandwidth requirements [5].

8.6.10 Leakage

One issue with CMOS sampling switches in submicron technologies, is that their open resistivity is not infinite [4]. Parasitic currents slowly discharge the voltage sampled on the capacitor. The effect is illustrated in Figure 8.20. Over time the stored voltage V_{store} droops by the voltage V_{droop}, hence producing an error on the voltage value that will be converted.

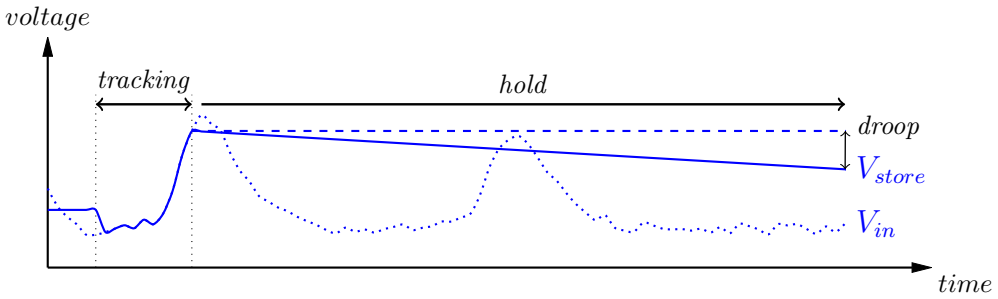


Figure 8.20 – Effect of leakage current on the sampled voltage V_{store} in a storage cell.

8.6.10.1 Model of switch leakage

Since the resistance *off* value, R_{off}, of the switch is very large when closed (in the $G\Omega$ range [4]), it creates an RC filter with a time constant much higher than the readout time. Therefore, it can be considered as a pure current source I_{off}, slowly discharging the storage capacitor C_S, as shown in Figure 8.21.

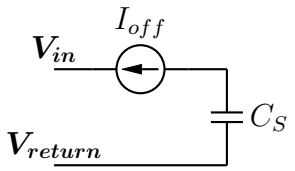


Figure 8.21 – Model of leakage current source in the channel.

Given enough time, the stored value, V_{store}, will slowly decrease towards the mean value of the input voltage.

8.6.10.2 Model for the NMOS transistor

To understand where the leakage current is coming from, we look at the current I_D going trough an NMOS transistor only for different v_{DS} voltage versus the applied *gate* voltage. The simulation results are reported in Figure 8.22. We can identify two regions for the current I_D:

Region one

$$V_G \ll V_S - V_T$$

In this region transistor is in *depletion region* [4], I_D results from the current flowing through the reverse *drain-bulk* diode. For an asymmetrical diode such as the *drain-bulk* diode, the reverse bias current I_S is given by:

$$I_S = qAn_i^2 \frac{D_P}{W_n N_D} \tag{8.4}$$

with, A the area of the diode, proportional to the switch size W; n_i, D_P, W_n, N_D geometrical and physical constants of the diode given by the technology used.

Region two

$$V_G \sim V_S - V_T$$

In this region current I_D starts to grow exponentially with the gate voltage V_G, the transistor enters the weak inversion region. In the weak inversion region, the *off* current of the NMOS transistor is given by [4]:

$$I_{off} = \frac{W}{L} I_{D0} exp\left(\frac{v_{GS}}{nkT/q}\right) \tag{8.5}$$

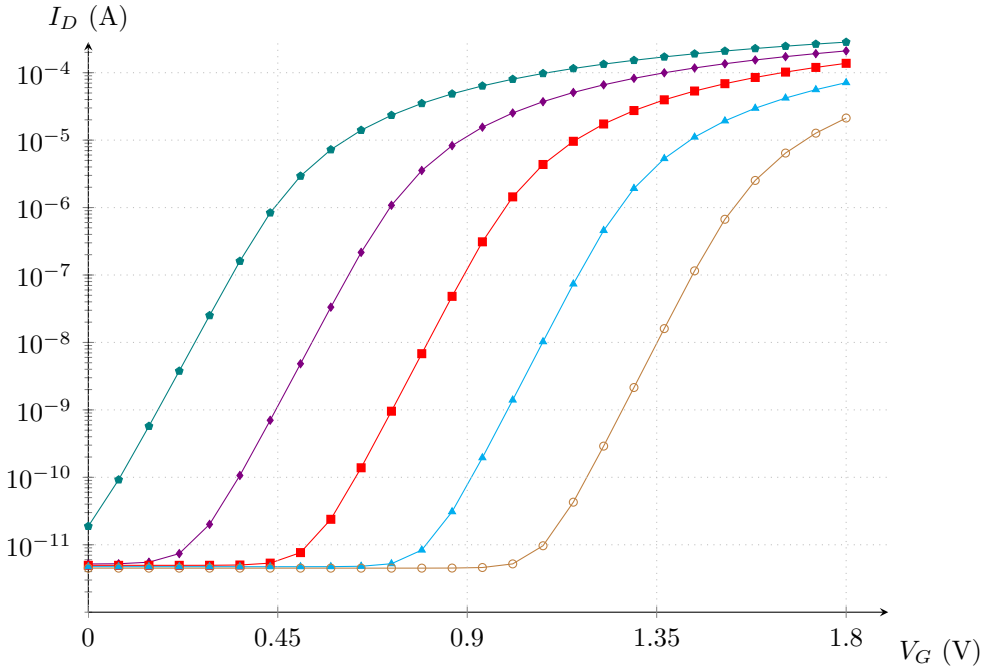
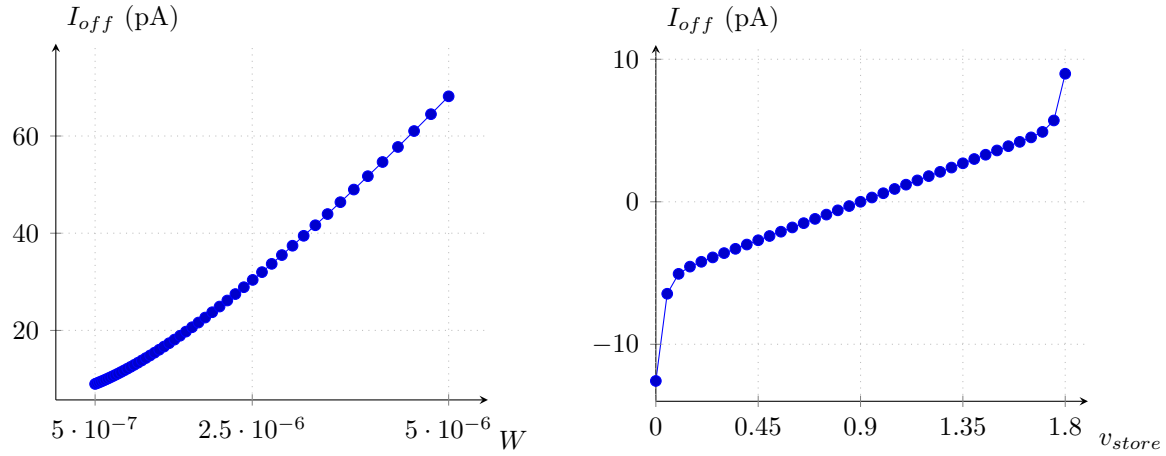


Figure 8.22 – Simulation of the current I_D through the switch versus the *gate* voltage V_G done for the drain voltage $V_D = 1.8V$ and for various values of source voltage V_S ($V_S = 0V$ —●—,$V_S = 0.23V$ —♦—,$V_S = 0.45V$ —■—,$V_S = 0.68V$ —▲—,$V_S = 0.9V$ —○—).

In order to reduce as much as possible the leakage current through the switch during *hold* mode, the weak inversion region, where I_D grows exponentially, is avoided and we stay where $V_G \ll V_S - V_T$.

8.6.10.3 Leakage in the switch

The switch used in the SamPic sampling cell is a CMOS switch made of a PMOS and a NMOS connected in parallel. The reason for this choice is to avoid threshold voltage limitation when using a single transistor only. A simulation of the $_{off}$ current versus the transistor size (W) is plotted in Figure 8.23a. We can verify the linear dependence of the I_{off} current as a function of the transistor size and note that, in this technology, the leakage current typically ranges from hundred's of fA, for small transistor sizes, to several ten's of pA.



(a) Simulation of the leakage current versus the transistor size W. We observe the linear dependence predicted by the I_{off} current formula.

(b) Simulation of the leakage current versus the v_{store} voltage on the switch, done for $v_{in} = 0.9$.

Figure 8.23 – Simulations of the leakage current versus the switch size W (a) and versus the v_{DS} of the switch (b).

Input voltage dependence Another characteristic of the switch off-current I_{off} is plotted in Figure 8.23b. The leakage is also a function of the voltage stored in the sampling capacitor. The dependence is a linear function of the voltage difference across the switch: $v_{DS} = V_{in} - V_{store}$, but grows exponentially when V_{store} reaches the voltage rails.

In the linear region, if we have A the coefficient of linearity, we can derive a simple model from Figure 8.23b:

$$I_{off} = A \times (V_{store} - V_{in}) \tag{8.6}$$

The equation that governs the discharge of the sampling capacitor C_S can be written as follows:

$$I_{off} = C_S \frac{dV_{store} - V_{in}}{dt}$$

We can consider that the variations of V_{in} over time (nanosecond range) are much faster than the discharge time of the sampling capacitor (millisecond range). So V_{store}

is only sensitive to V_{in} mean voltage $V_{ref} = \lceil V_{in} \rceil$. Hence, if $V_{store}(t_0)$ is the sampled voltage on the capacitor, the voltage on the sampling capacitor after the time t_{hold} is given by:

$$V_{store}(t_{hold}) = V_{ref} + [V_{store}(t_0) - V_{ref}]\, e^{-\frac{A}{C_S}(t_{hold}-t_0)} \tag{8.7}$$

So the relative error $\epsilon(t_{hold})$ due to leakage made on the read voltage V_{store} after the time t_{hold} can be written as:

$$\epsilon(t_{hold}) = 1 - \frac{V_{store}(t_{hold}) - V_{ref}}{V_{store}(t_0) - V_{ref}} = 1 - e^{-\frac{A}{C_S}(t_{hold}-t_0)} \tag{8.8}$$

The characteristic time constant of the discharge t_{dis} can be identified to $t_{dis} = \frac{C_S}{A}$.

$$t_{dis} = 6.3ms \tag{8.9}$$

If we take $C_S = 35$ fF and $A = \frac{5pA}{0.9V} = 5.6 \times 10^{-12}\Omega^{-1}$ we have:

Table 8.2 – Values of the error ϵ due to the leackage on the storage cell for various values of time delay $t_{hold} - t_0$.

$t_{hold} - t_0$	error ϵ
500 ns	8.0×10^{-5}
1 μs	2.0×10^{-4}
2 μs	3.2×10^{-4}
5 μs	8.0×10^{-4}

Since typically $t_{hold} - t_0 \ll t_{dis}$, the error ϵ due to leakage grows linearly with the *hold* time t_{hold}.

8.6.10.4 Effect of the leakage on sampling

The effect of the switch leakage current on sampled pulses is illustrated in Figure 8.24.

As described previously, this effect is mostly linear except close to the rail where the leakage effects are much stronger, and can be seen as a saturation effect on the sampled pulse after a short period of time. Since the effects of *leakage* are determinist, they can be calibrated and corrected after digital conversion and chip readout.

Temperature dependence The simulation of the temperature dependence of the leakage voltage I_{off} versus the stored voltage V_{store} is shown in Figure 8.25. In this Figure we can see two interesting features:

- The slope A of I_{off} with V_{store} is not temperature-dependent. Meaning that for a given switch the *discharging* time t_{dis} will not change with the temperature

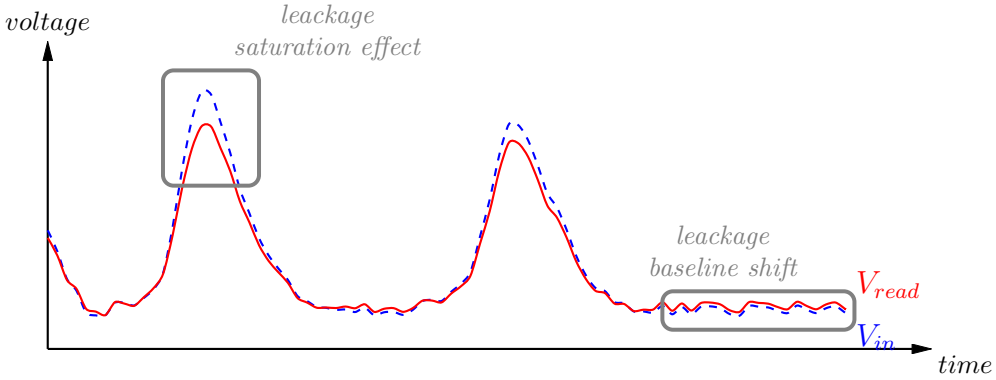


Figure 8.24 – Illustration of the leakage effects on the sampled signal (red) compared to the original signal (dashed blue): near the rails the leakage has an exponential dependence and can be seen as a saturation effect.

- However, the range where $I_{off} = A \times (V_{store} - V_{in})$ shrinks with increasing temperature. So at higher temperature, the *saturation* effects due to exponential leakage near the rails will increase and reduce the input dynamic by up to 200 mV at $T = 100°C$.

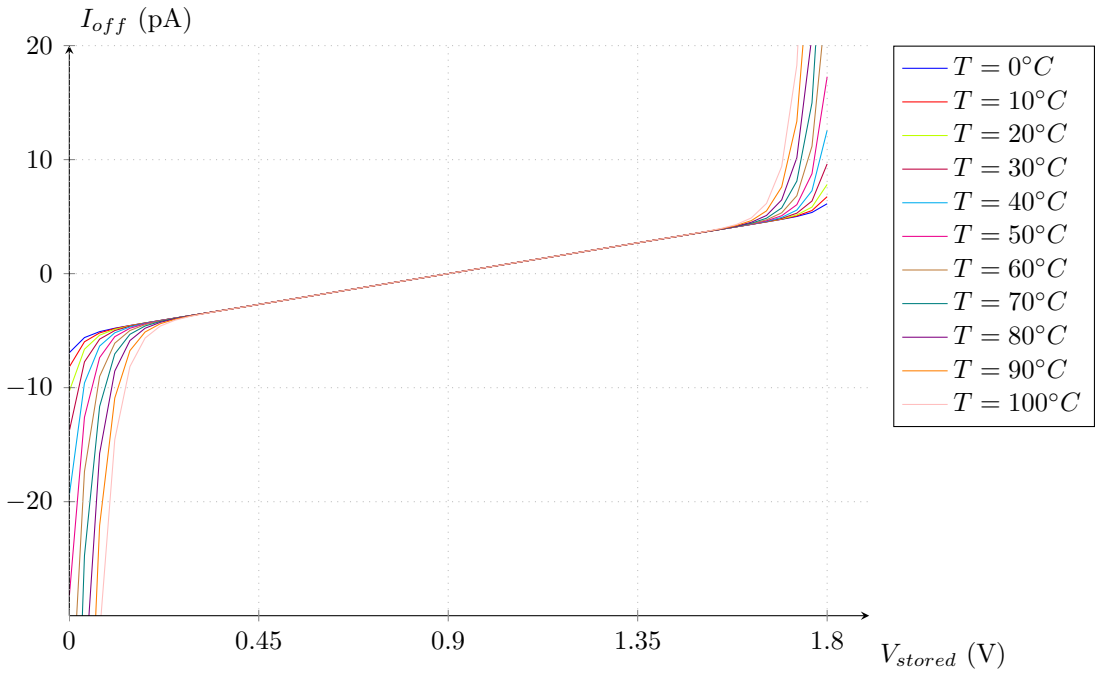


Figure 8.25 – Simulation of the leakage current versus the stored voltage V_{store} on the switch, for a wide range of temperature variations (from $T = 0°C$ to $T = 100°C$).

8.6.11 Bandwidth

When closed, the switch resistance value is R_{ON}, creating a RC lowpass filter with the sampling capacitor C_S (see Figure 8.26).

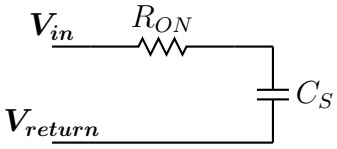


Figure 8.26 – Equivalent circuit of the sampling cell when the switch is closed.

The ON resistance formula of a PMOS and NMOS switch with size W and L is given by [4]:

$$R_{ON} = \frac{1}{\frac{W}{L}(\mu_n C_{ox}(V_{DD} - V_{in} - V_{THN}))(\mu_p C_{ox}(V_{in} - |V_{THP}|))} \tag{8.10}$$

The simulation of the switch ON resistance as a function of the input voltage V_{in} for different transistor sizes is plotted in Figure 8.27.

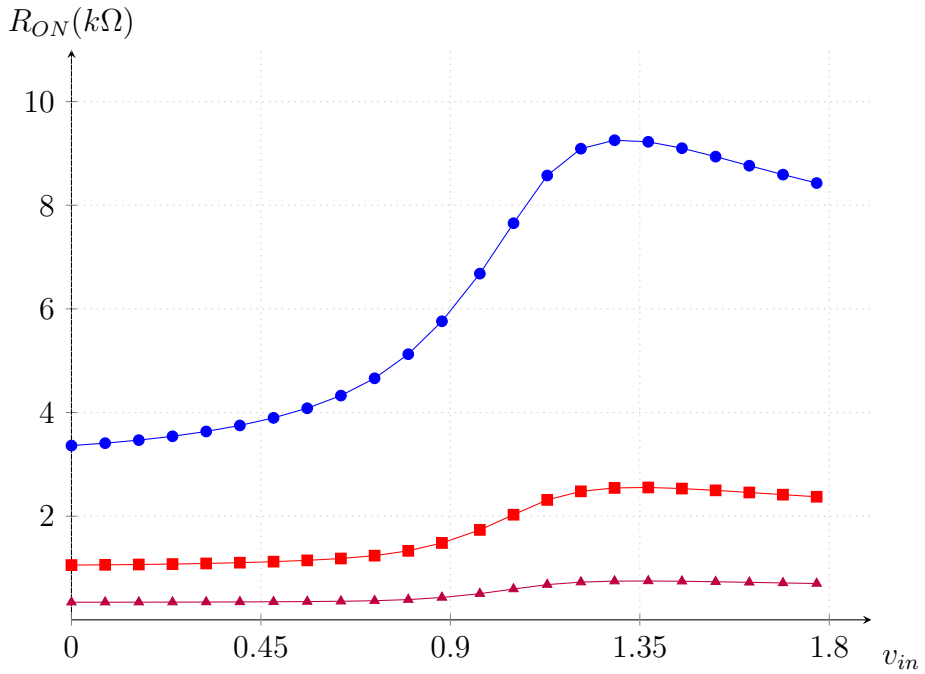


Figure 8.27 – Simulation of the ON resistance of the switch as a function of the input voltage v_{in} for three different switch size: 500nm —●—, 1.6μm —■— and 5μm —▲—. The switch PMOS and NMOS sizes are equal.

The 3 dB cutoff frequency of the filter f_{max} made of a sampling capacitor C_S and the switch is given by:

$$f_{max} = \frac{1}{2\pi R_{ON} C_S} \tag{8.11}$$

Above this frequency, less than half the power of the input signal couples to the sampling capacitor. We can make three observations:

- In order to record very fast signals, a high bandwidth is required, hence requiring to reduce either the sampling capacitor C_S value or the ON resistance of the switch R_{ON}.
- The ON resistance of the switch is not a steady function of the voltage across it, and R_{ON} decreases with larger transistors. The switch dimensions must be adjusted to have enough bandwidth over the full range of the input signal V_{in} and reduce as much as possible the non-linearity effects of the ON resistance.
- Simultaneously, the switch resistance R_{off} must be kept high to avoid important leakage in the *off* state.

An AC simulation of the switch attenuation effect on the sampled voltages in the sampling cell is plotted in Figure 8.28. For a sampling cell of size $W = 2\ \mu m$ and $C_S = 35$ fF the cut-off frequency is $f_c = 4.8$ GHz.

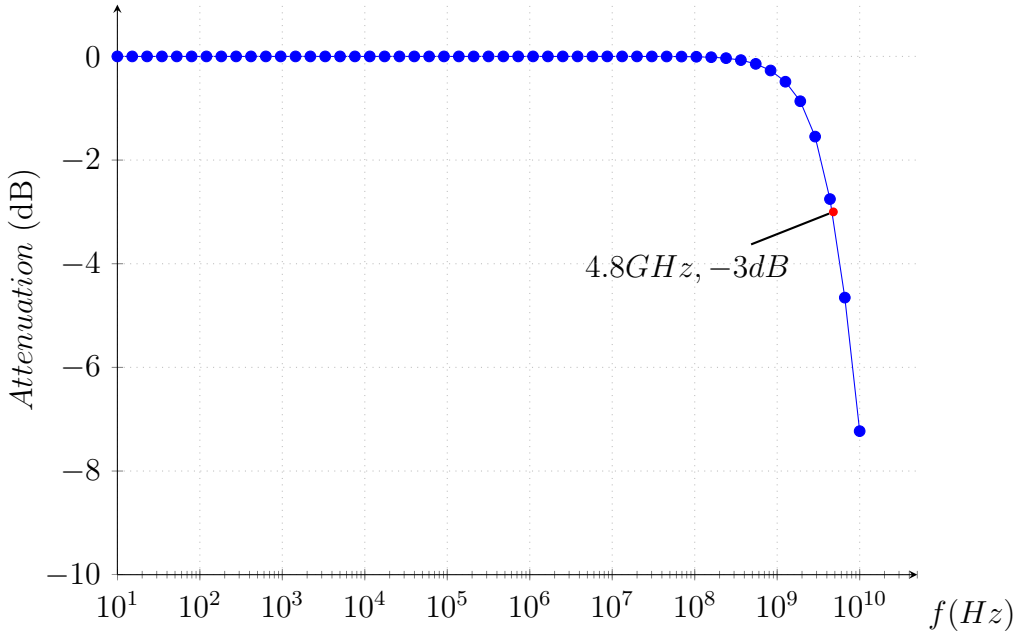


Figure 8.28 – Simulation of the minimum bandwidth of the sampling cell versus the frequency ($W = 2\mu m$, $C_S = 35fF$).

8.6.11.1 Bandwidth limitation errors

The effect of bandwidth limitations is illustrated in Figure 8.29. For input signals with frequency spectrum ranges close to the *cutoff* frequency f_{max} of the sampling cell, gain errors are made on fast signals. This effect, although deterministic, is complicated to correct for, as the corrections are frequency and amplitude dependent. Other illustrations of this effect can be found in Figure 8.36.

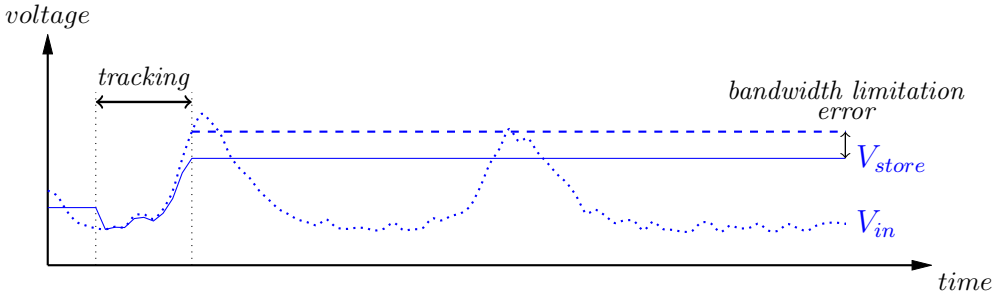
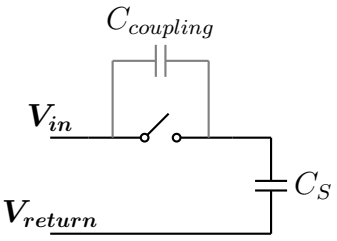


Figure 8.29 – Illustration of the bandwidth limitation effect on the sampled signal. Due to bandwidth limitations, the voltage on the sampling capacitor cannot follow a very fast input signal, yielding to gain error for fast signals.

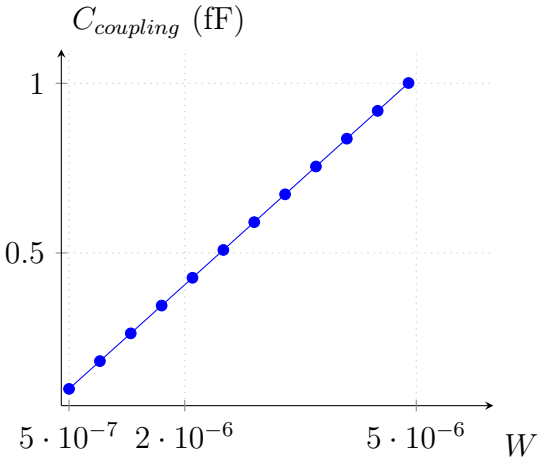
8.6.12 Input coupling

A parasitic capacitor C_{DS} or $C_{coupling}$ is present between the input node V_{in} and the storage node V_{store} (see Figure 8.30a). This parasitic capacitor allows fast AC input signal to couple to the *stored* signal and degrade it.

The typical value for $C_{coupling}$ are plotted in Figure 8.30b as a function of the switch size. A small switch results in less coupling (smaller $C_{coupling}$ value).



(a) Schematic model of the sampling circuit showing the parasitic capacitor in parallel with the sampling switch.



(b) Simulation of the value of the parasitic capacitor $C_{coupling}$ of the sampling switch as a function of its size W.

Figure 8.30 – Description of the input coupling into the sampling capacitor.

At very high frequency, the circuit of Figure 8.30a can be seen as a capacitive voltage divider and we compute:

$$V_{C_S} = V_{store} + V_{in}\frac{C_{coupling}}{C_S + C_{coupling}} \tag{8.12}$$

Typically the coupling capacitor C_{DS} or $C_{coupling}$ is very small (see Fig. 8.30b), but for high speed circuit applications, where the input signal can range up to the

GHz and with very small sampling capacitor value C_S, this effect occurs at the few percent level.

For example, the calculation of the maximum percentage of coupling with $C_S = 35$ fF is given in the Table below:

Table 8.3 – Maximum percentage of coupling with $C_S = 35$ fF.

V_{in}	$C_{coupling} = 0.5$ fF	$C_{coupling} = 1$ fF
0.9 V	1.4%	2.7%

8.6.12.1 Input coupling noise effect

The illustration of the coupling effect of the input voltage V_{in} onto the stored voltage V_{store} is represented in Figure 8.31. Since this effect depends of the variation of V_{in} through time, it cannot be calibrated out or filtered and must be considered as additional noise on the sampled signal V_{store}.

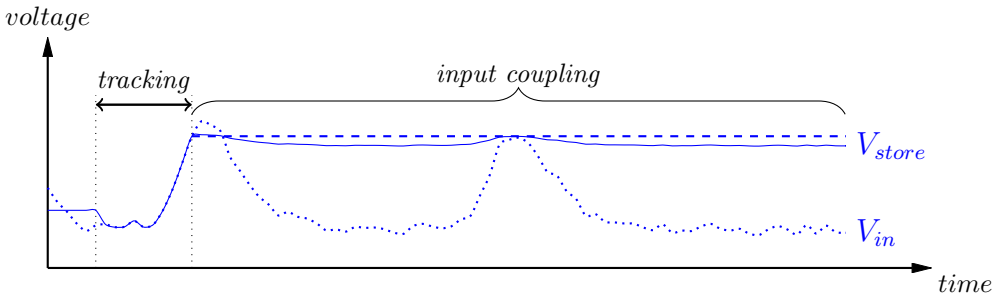


Figure 8.31 – Illustration of the coupling effect of the input voltage V_{in} onto the stored voltage V_{store}. Since this effect depends of the value of V_{in}, through time it cannot be calibrated out or filtered and must be considered as additional noise on the sampled signal V_{store}.

Spectral representation The AC simulation of the input coupling effect is plotted in Figure 8.32. The coupling of input signals with frequencies ranging from 1 Hz to 10^{11} Hz has been simulated for a sampling cell with and without the extracted parasitic capacitances inside the cell. Due to the complexity of the sampling cell layout (see Figure 8.53), it is difficult to model and predict couplings to high frequencies signals inside the cell. Simple models are wrong most of time and simulation with parasitic effects are imperative.

At low frequency, the coupling goes to $\sim$ 0dB, which is simply illustrating the effect of leakage through the switch. Overall, we observe that the pollution due to coupling through the sampling switch contributes several orders of magnitude more that the simple kTC noise of the sampling capacitor. This issue is addressed using a different sampling cell architecture (see Figure 8.48).

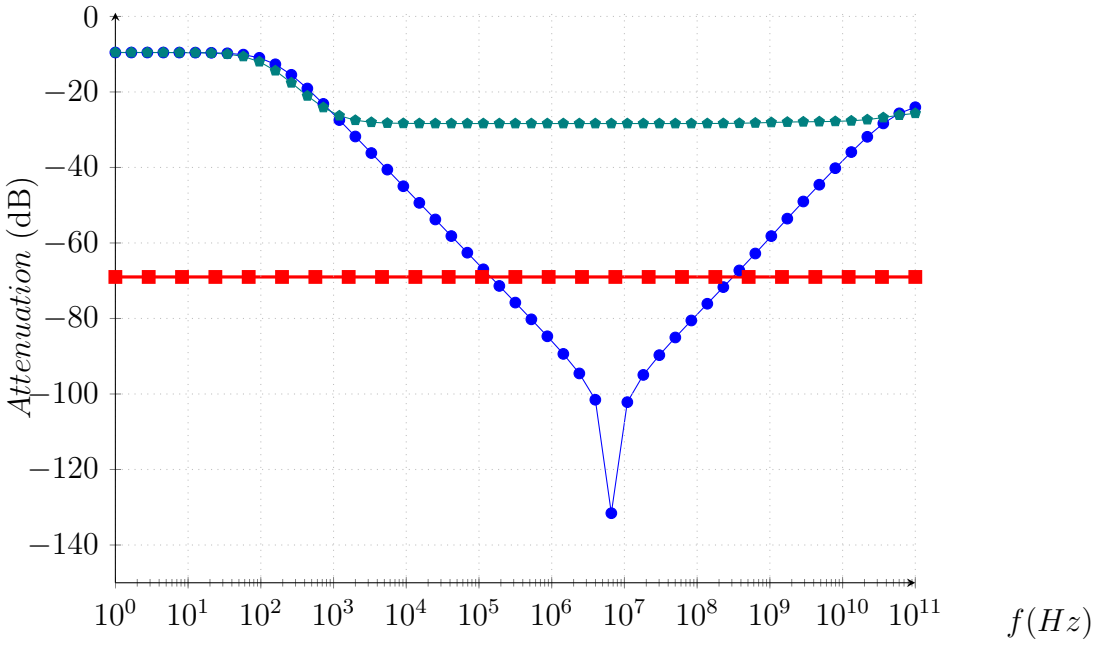


Figure 8.32 – AC simulation of the coupling magnitude from the input signal V_{in} to the sampling capacitor, in dB. The simulation was done for coupling with and without considering parasitic effect of the layout design. The kTC noise of the sampling cell is also printed for comparison.

8.6.13 Charge injection

Considering only an NMOS transistor used as a switch (Figure 8.33), when the switch in ON some charge Q_{ch} is stored in the channel. The total charge can be modeled as[4]:

$$Q_{ch} = WLC_{ox}(V_{DD} - V_{in} - V_{TH}) \tag{8.13}$$

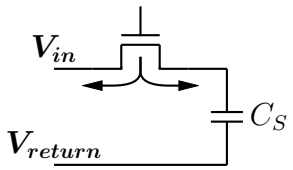


Figure 8.33 – Charge injection when the switch turns off, illustrated with an NMOS switch.

When the switch is opened, the charge stored Q_{ch} exits through the drain and source of the transistor. If we consider that half of the charge it sent to the right on C_S (Figure 8.33), the resulting voltage offset on C_S is:

$$\Delta V = \frac{WLC_{ox}(V_{DD} - V_{in} - V_{TH})}{2C_S} \tag{8.14}$$

The charge injection effect is illustrated in Figure 8.34. At the time of the sampling, the charge injected on the sampling capacitor C_S creates an offset on the

sampled voltage V_{store}.

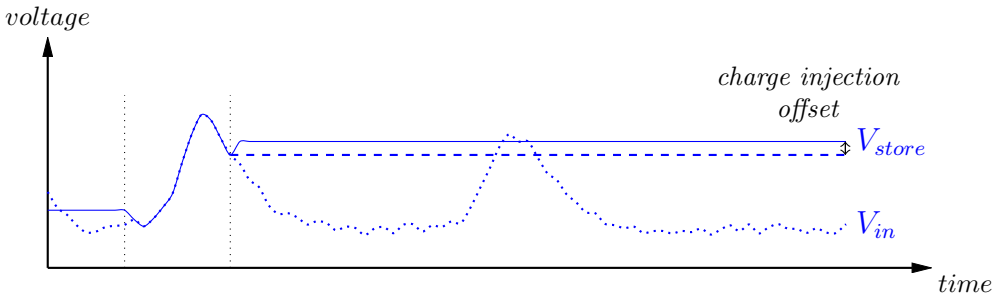


Figure 8.34 – Effect of the charge injection during sampling. The injected charge creates an offset on the sampled voltage V_{store}

Taking into account the body effect of the NMOS: $V_{TH} = V_{TH0} + \gamma(\sqrt{2\phi_B + V_{in}} - \sqrt{2\phi_B})$ it follows that the error ΔV exhibits a non-linear dependence upon V_{in}. This non-linear effect increases with the size of the switch, as illustrated in Figure 8.35.

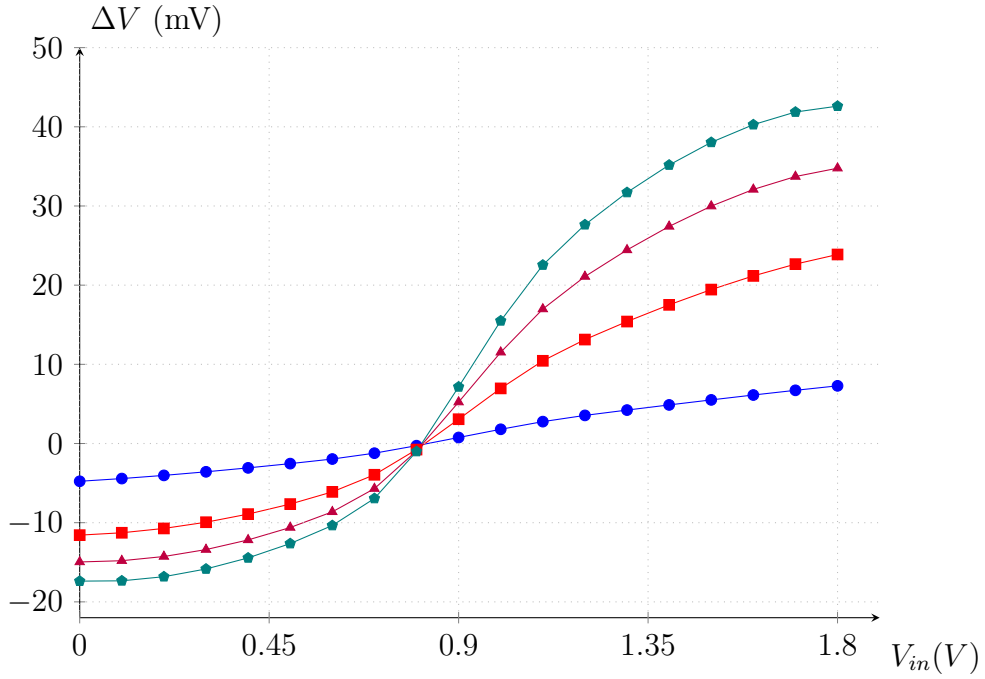


Figure 8.35 – Simualtion of the voltage offset ΔV as a function of the input voltage V_{in} for different size W ($W = 500nm$ —●—,$W = 2\mu m$ —■—, $W = 3.5\mu m$ —▲—, $W = 5\mu m$ —⬟—) of the sampling switch. Simulation is done with a sampling capacitor value $C_S = 35fF$.

It is interesting to consider the speed-to-precision trade-off resulting from the charge injection. If we represent the sampling speed by a time constant τ and the precision by the error ΔV due to the charge injection, we can define a figure of merit $F = {}^1/_{\tau\Delta V}$ [4]. Considering only an NMOS switch

$$\tau = R_{ON}C_S = \frac{1}{\mu_n C_{ox}(W/L)(V_{DD} - V_{in} - V_{TH})}C_S \tag{8.15}$$

and from Equation 8.14 we have:

$$F = \frac{\mu_n}{2L^2} \tag{8.16}$$

Thus, to the first order, the trade-off is independent of the switch width and the sampling capacitor. Once the bandwidth of the sampling cell is set, we have no more freedom to decrease the non-linearity due to the charge injection.

Therefore the bandwidth of the cell must be chosen just high enough to accomodate for the bandwidth of the signal of interest (via W size optimization), but not higher. This to avoid degradation of the signal due to the non-linear effect of charge injection for large W sizes.

Charge injection corrections Charge injection is a deterministic effect, however, as can be seen in Figure 8.35, it is a non-linear effect. Corrections require third order spline models or lookup tables, which makes them computationally expensive to use in real time data acquisition systems.

8.6.14 Residual charge

During their tracking phase, the sampling cells are reconnected to the input bus. If we consider that the previous voltage value on the capacitor was V_{store} and $\tau = R_{ON}C_S$, the voltage on the capacitor follows the differential equation:

$$\tau \frac{dV_{cap}}{dt} + V_{cap}(t) = V_{in}(t) \tag{8.17}$$

Assuming that at $t = t_0$ the switch is closed we can write the solution:

$$V_{cap}(t) = e^{-(t-t_0)/\tau}\left(V_{store} + \int_{t_0}^{t} \frac{V_{in}(t)}{\tau} e^{(t-t_0)/\tau} dt\right) \tag{8.18}$$

Which expands into:

$$V_{cap}(t) = \underbrace{V_{in}(t)}_{\text{Tracking}} + \underbrace{V_{store}e^{-(t-t_0)/\tau}}_{\text{Discharging}} - \underbrace{V_{in}(t_0)e^{-(t-t_0)/\tau}}_{\text{Charging}} - \underbrace{e^{-(t-t_0)/\tau}\int_{t_0}^{t} \frac{dV_{in}(t)}{dt} e^{(t-t_0)/\tau} dt}_{\text{Bandwidth}}$$

The solution consists of four terms as listed in the expanded form of Equation 8.18.

- A *tracking* term $V_{in}(t)$: in established mode the voltage on the capacitor is tracking the input voltage.
- A *discharging* term $V_{store}e^{-(t-t_0)/\tau}$: due to the transient phase during which the previously stored voltage is discharged on the input bus.
- A *charging* term $V_{in}(t_0)e^{-(t-t_0)/\tau}$: due to the transient phase during which the capacitor charges up to the initial input voltage.
- A *bandwidth* term $e^{-(t-t_0)/\tau}\int_{t_0}^{t} \frac{dV_{in}(t)}{dt} e^{(t-t_0)/\tau} dt$.

To evaluate the error and justify the name given to the fourth term: $e^{-(t-t_0)/\tau}\int_{t_0}^{t}\frac{dV_{in}(t)}{dt}e^{t-t_0/\tau}dt$, lets consider an input signal V_{in} extending over the finite frequency range from $f=0$ to f_{max}. We can write the Fourier decomposition of the signal:

$$V_{in}(t)=\Re\left(\int_0^{f_{max}}A(f)e^{2i\pi ft}df\right) \tag{8.19}$$

And:

$$\frac{dVin(t)}{dt}=\Re\left(\int_0^{f_{max}}2iA(f)\pi fe^{2i\pi ft}df\right)$$

In the case where we consider an input signal V_{in} made of a single frequency f_{in} of amplitude A: $V_{in}(t)=\Re\left(Ae^{2i\pi f_{in}t}\right)$, the integration of the error term becomes simple and we can write:

$$\begin{aligned}e^{-(t-t_0)/\tau}\int_{t_0}^{t}\frac{dV_{in}(t)}{dt}e^{t-t_0/\tau}dt&=\Re\left(2iA\pi f_{in}e^{-(t-t_0)/\tau}\int_{t_0}^{t}e^{2i\pi f_{in}t}\times e^{(t-t_0)/\tau}dt\right)\\&=\Re\left(2iA\pi f_{in}e^{-(t-t_0)/\tau}\left[\frac{e^{2i\pi f_{in}t}\times e^{(t-t_0)/\tau}}{2i\pi f_{in}+\frac{1}{\tau}}\right]_{t_0}^{t}\right)\\&=\Re\left(2iA\pi f_{in}\frac{e^{2i\pi f_{in}t}-e^{2i\pi f_{in}t_0}e^{-(t-t_0)/\tau}}{2i\pi f_{in}+\frac{1}{\tau}}\right)\end{aligned}$$

If we write:

$$\frac{1}{\tau}=2\pi f_c$$

Two terms can be identified. The first one is a transient:

$$\Re\left(Ae^{2i\pi f_{in}t_0}\frac{if_{in}}{f_c+if_{in}}\right)e^{-(t-t_0)/\tau}\leq V_{in}(t_0)\frac{f_{in}}{\sqrt{f_c^2+f_{in}^2}}e^{-(t-t_0)/\tau} \tag{8.20}$$

and second one is:

$$\Re\left(Ae^{2i\pi f_{in}t}\frac{if_{in}}{f_c+if_{in}}\right)\leq V_{in}(t)\frac{f_{in}}{\sqrt{f_c^2+f_{in}^2}} \tag{8.21}$$

We can identify these error terms to the well known attenuation due to the cutoff frequency f_c of the $R_{ON}C_S$ low-pass filter.

The phase shift can be calculated to be:

$$\Delta\phi=\frac{\pi}{2}-tan^{-1}\left(\frac{f_{in}}{f_c}\right) \tag{8.22}$$

Charging and discharging errors We calculate the charging and discharging error assuming $f_c=1$ GHz and the maximum voltage swing of 1.8 V.

Table 8.4 – Errors due to the charging and discharging transients on the sampling capacitor for $f_c = 1$ GHz and a maximum voltage swing of 1.8 V.

$t - t_0$ in ps	Error in mV
100	960.28
200	512.30
400	145.80
800	11.81
1600	0.08

Table 8.5 – Errors due to the bandwidth limitation for $f_c = 1$ GHz.

f_{in}	Error in mV
5 MHz	0.02
10 MHz	0.10
50 MHz	2.49
100 MHz	9.90
500 MHz	200.00
1 GHz	500.00

Bandwidth errors We calculate the error (distortion) due to the bandwidth limitation assuming $f_c = 1$ GHz and an input signal of amplitude 1.8 V.
In order to reduce these errors we can tune two parameters:

- The tracking time $t - t_0$, which allows exponential reduction of all the transient errors. However a longer tracking time implies more cells connected simultaneously to the input bus
- The bandwidth of the sampling cell f_c, which allows exponential reduction of transient errors as well and the frequency distortions as well. However increasing the bandwidth will increase the distortion due to charge injection.

8.6.14.1 Basic simulation

A simulation of the charging effects on the sampled signal inside SamPic is shown in Figure 8.36. In each case, we simulated the deformation effect of a Gaussian pulse due to the limited tracking time during sampling. Since this effect can be assimilated to a bandwidth limitation, three effects can be observed:

- **Signal attenuation** When the tracking time is too short, the signal cannot fully develop in the sampling cell and is attenuated. If the tracking time is not well defined in time or varies from cell to cell, this can be seen as a noise effect.
- **Delay or phase shift** When the tracking time is too short, a phase shift occurs in the signal. This effect is more problematic if it isn't deterministic, since it will create large errors when we will try to measure the time of the signal.

- **Ghost pulse** During the simulation we start from a state where a pulse was stored around cell number 15 (see Figure 8.36). In some cases, where the tracking time is too short, we can observe a "ghost" of the previous pulse.

The *internal* bandwidth of the sampling cell in SamPic (switch + capacitor) is designed to be large enough to mitigate these undesirable effects.

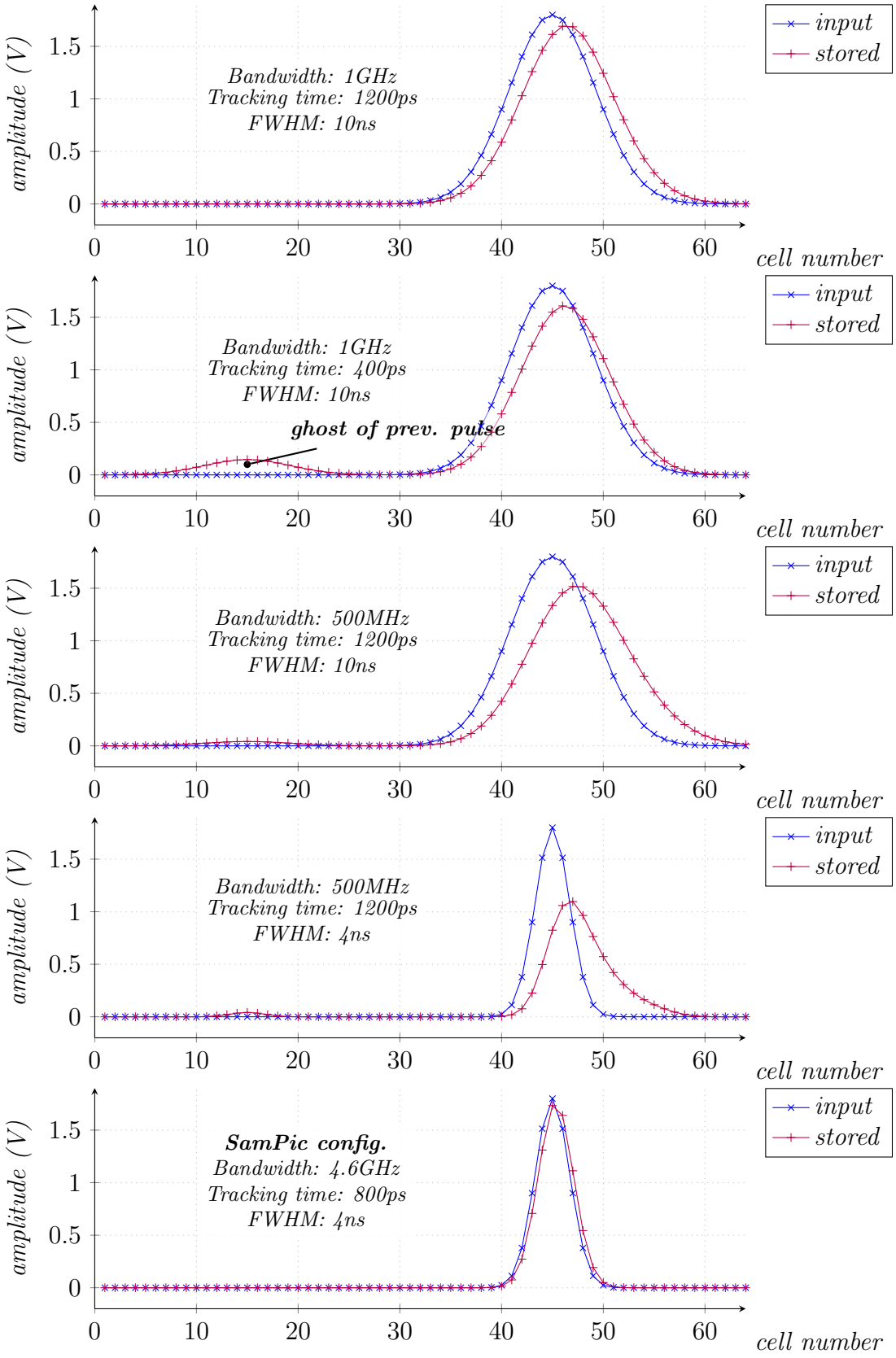


Figure 8.36 – Simulation of Gaussian pulses sampling for various values of input bandwidth, tracking time and pulse width at 1 GSa/s.

8.6.15 Ghost pulses

Another perturbative effect happens when connecting multiple sampling cells to one common input bus. That effect is illustrated in Figure 8.37.

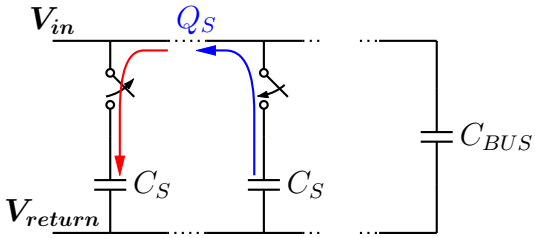


Figure 8.37 – Illustration of the ghost pulse creation by charge transfer. At each sampling time step in the channel, we have simultaneously one cell discharging (blue arrow) and one cell charging (red arrow). Due to the non-infinite time response of the input bus, some of the released charge Q_S is resampled.

During each sampling step, one cell that was previously in *hold* mode is reconnected to the input bus and goes in *track* mode (see Fig. 8.16). If the cell was holding the voltage V_{prev}, it releases the charge $Q_S = C_S(V_{in} - V_{prev})$ on the input bus. With C_{BUS} the capacitor between V_{in} and V_{bias}, the voltage offset created by the charge released on the bus is given by:

$$\Delta U = (V_{in} - V_{prev}) \frac{C_S}{C_{BUS}} \tag{8.23}$$

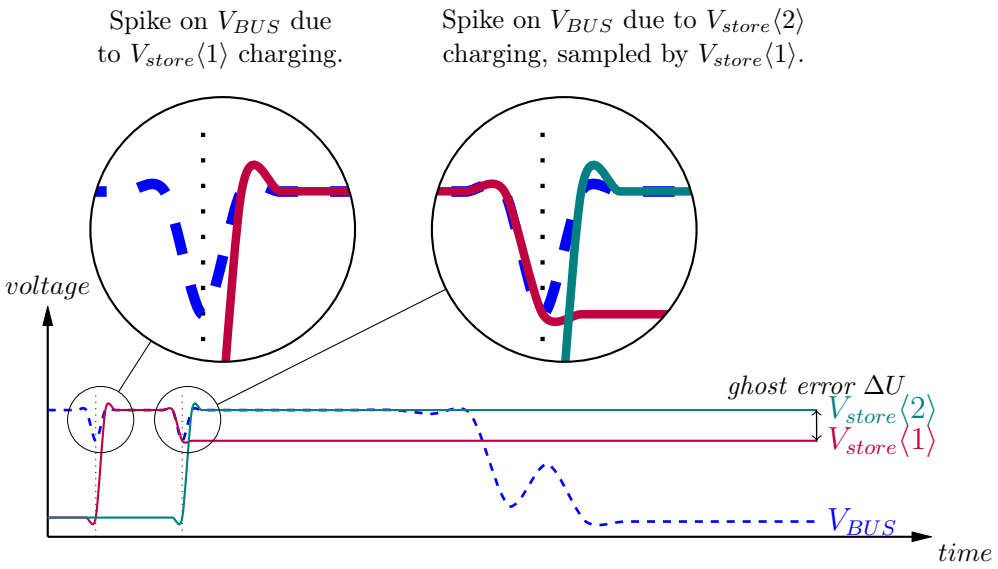


Figure 8.38 – Illustration of ghost creation in a sampling capacitor array. The voltage ΔU created by the *discharge* of the sampling cell 2 ($V_{store}\langle 2\rangle$), is sampled by $V_{store}\langle 1\rangle$.

When simultaneously to the charge release, one cell goes from *track* mode to *hold* mode, it samples $V_{in} + \Delta U$. Because the perturbation ΔU depends on the

previously sampled signal V_{prev}, it can be seen as a ghost pulse in the sampling array. Interestingly, the ghost pulse is an image of the previous pulse shifted by the number of cells in the tracking window. The principle is illustrated in Figure 8.38.

The effect of the perturbation ΔU on the input line, as a function of the previous voltage V_{prev}, for $V_{in} = 0.9$ V is plotted in Figure 8.39. Due to the complexity of this process, which is a function of many parameters such as: the parasitic capacitor of the bus C_{BUS}, the sampling time (see Figure 8.38), the input impedance, etc., the simulation results reported are only indicative. The order of magnitude of this perturbation is about 1 mV and is strongly dependent on the previous state of the sampling capacitor V_{prev}.

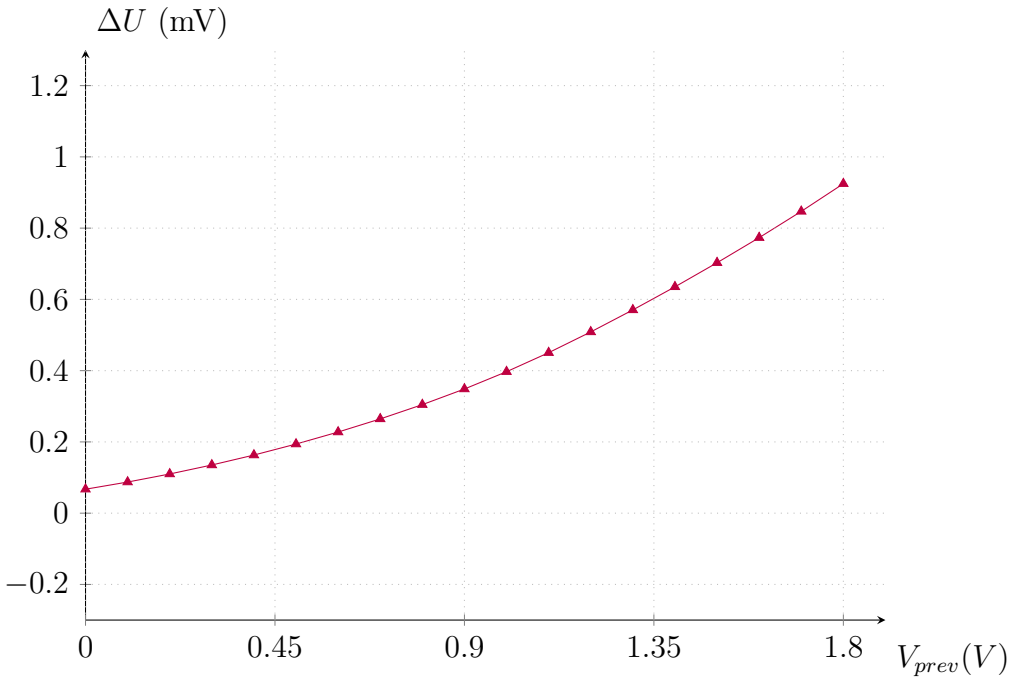


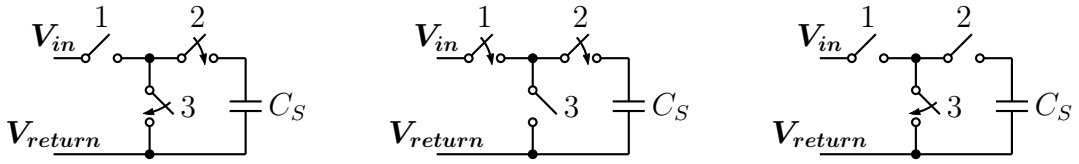
Figure 8.39 – Simulation of the discharge effect illustrated in Figure 8.37. The error voltage ΔU —▲— function of the value of V_{prev} is shown for $V_{in} = 0.9V$.

After injection the offset voltage is discharged through the input resistance $R_{ON} = 50\Omega$ of the input channel.

$$\Delta U = (V_{in} - V_{prev}) \frac{C_S}{C_{BUS}} e^{-(t - t_0)/R_{ON} C_{BUS}} \tag{8.24}$$

8.6.16 Sampling cell with reset

An improvement of this sampling cell is shown in Figure 8.40. Two switches are added to the previous cell in order to be able to perform a reset of the sampling cell before the tracking.



(a) Reset stage of the cell. (b) Sampling stage of the cell. (c) Hold stage of the cell.

Figure 8.40 – Sampling stages of the cell. During *reset*, the switches 2 and 3 are closed, 1 is opened. During *sampling* the switches 1 is closed and 3 opened. During *hold*, switches 1 and 2 are opened and 3 is closed.

The sampling sequence of the new cell is shown in Figure 8.40 and illustrated in Figure 8.41. It breaks down into three steps, described in the next subsections.

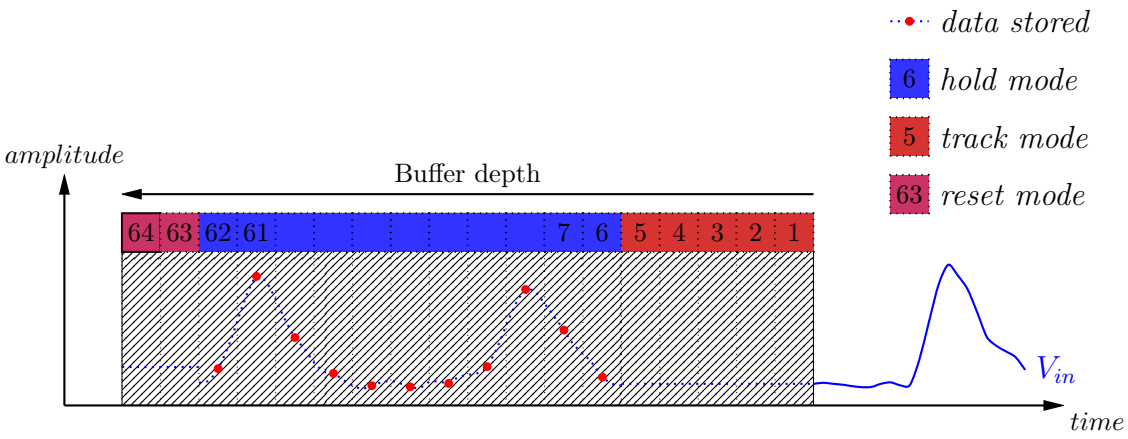


Figure 8.41 – Principle of sampling in the channel with the sampling cell of Figure 8.40. In addition to the track and hold stages described in Figure 8.16, a further stage is added during which the sampling cells are reset to a constant, known voltage.

8.6.16.1 Reset stage

During this stage the switches 2 and 3 of the sampling cell (Figure 8.41) are closed. Switch 1 stays open. The sampling capacitor C_S is shorted and discharged to the value V_{return}. (Figure 8.40a).

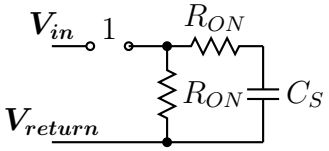


Figure 8.42 – Illustration of the reset stage: switches 2 and 3 are closed and switch 1 is open. The capacitor C_S is discharged to V_{return}.

The main advantage of this additional step is to reset all the charges of the sampling capacitor to a known value, thus removing the charge release errors depicted in Figure 8.37 and 8.39.

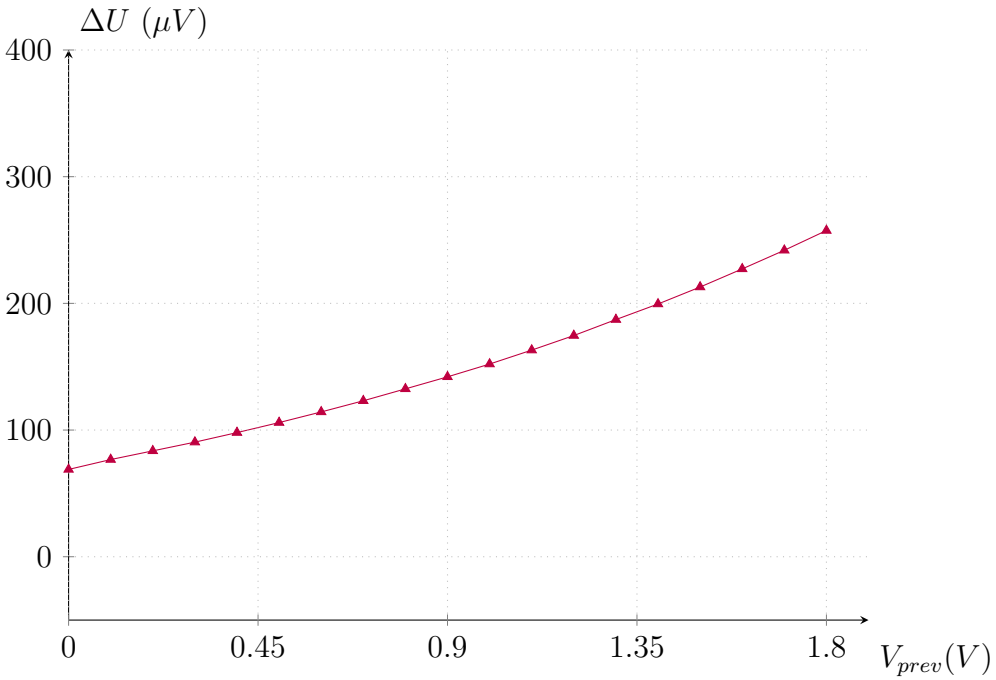


Figure 8.43 – Simulation of the discharge effect illustrated in Figure 8.37. The error voltage ΔU —▲— versus the value of V_{prev} for $V_{in} = 0.9V$. We have an improvement by a factor 10 compared to Figure 8.39.

This step is performed for each cell after being in *hold* mode and before releasing the sampling cell into *track* mode, see illustration in Figure 8.41.

8.6.16.2 Sampling stage

During this stage the switches 1 and 2 are closed and switch 3 is open. The capacitor C_S is connected between V_{in} and V_{return} and charges up (Figure 8.40b).

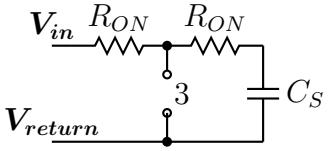


Figure 8.44 – Illustration of the sampling stage in the sampling cell with reset. The operation is similar to a standard sampling cell, switches one and two are closed and three is open. The R_{ON} resistance in this configuration is doubled.

A disadvantage of this structure is that we have two switches in series to charge C_S, which yields to an ON resistance multiplied by 2. Thus, the bandwidth is divided by two while the error due to charge injection stays unchanged.

Another effect is illustrated in Figure 8.45. Due to the non-linear dependence of the ON resistance of the switch, the risetime of the voltage on the capacitor is a function of the reset voltage of the switch V_{reset}. Setting the *reset* voltage of the switch to a low value allows faster risetimes on average.

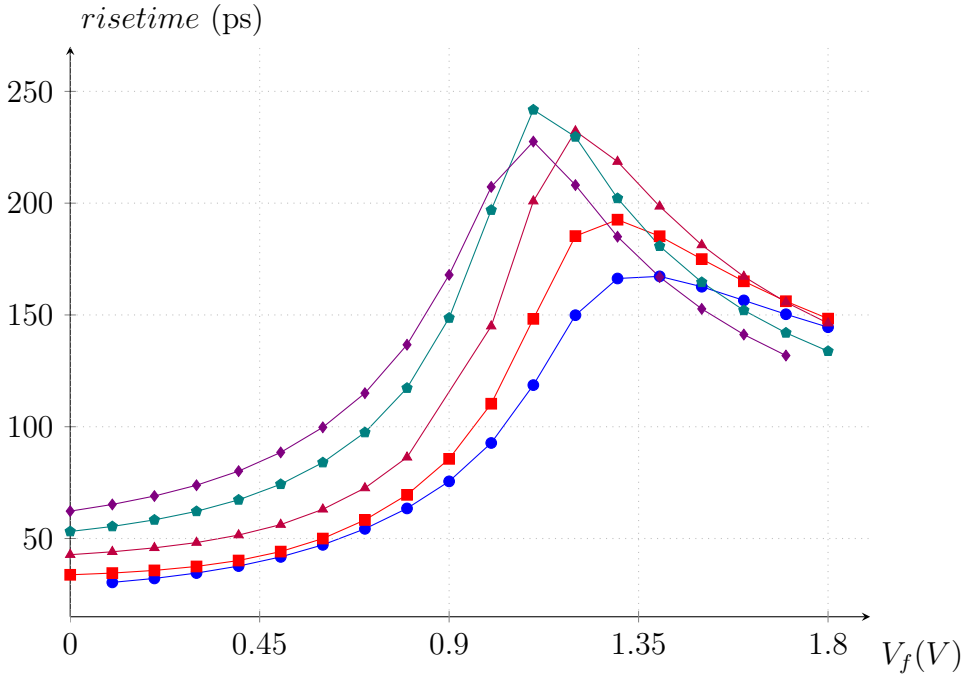


Figure 8.45 – Response time (risetime from 10% to 90%) of the sampling capacitor, for an input step going from V_{reset} to V_f for $V_{reset} = 0V$ —●—, $V_{reset} = 0.45V$ —■—, $V_{reset} = 0.9V$ —▲—, $V_{reset} = 1.35$ —⬟—, $V_{reset} = 1.8V$ —◆—.

8.6.16.3 Hold stage

During this stage the switches 1 and 2 are open and the switch 3 is closed. The capacitor C_S holds its value. (Figure 8.40c).

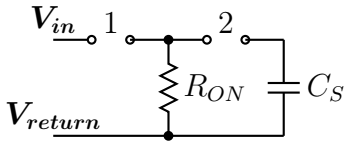


Figure 8.46 – Illustration of the hold stage in the sampling cell with reset. The principle difference to a standard sampling cell is the isolation between the sampling capacitor C_S and the input line V_{in}, decoupling the two nodes.

Leakage control By controlling the *reset* voltage applied on the internal node I during the *hold* period, it is possible to control the maximum value for v_{DS} to avoid staying in region of large leakage current (Figure 8.47).

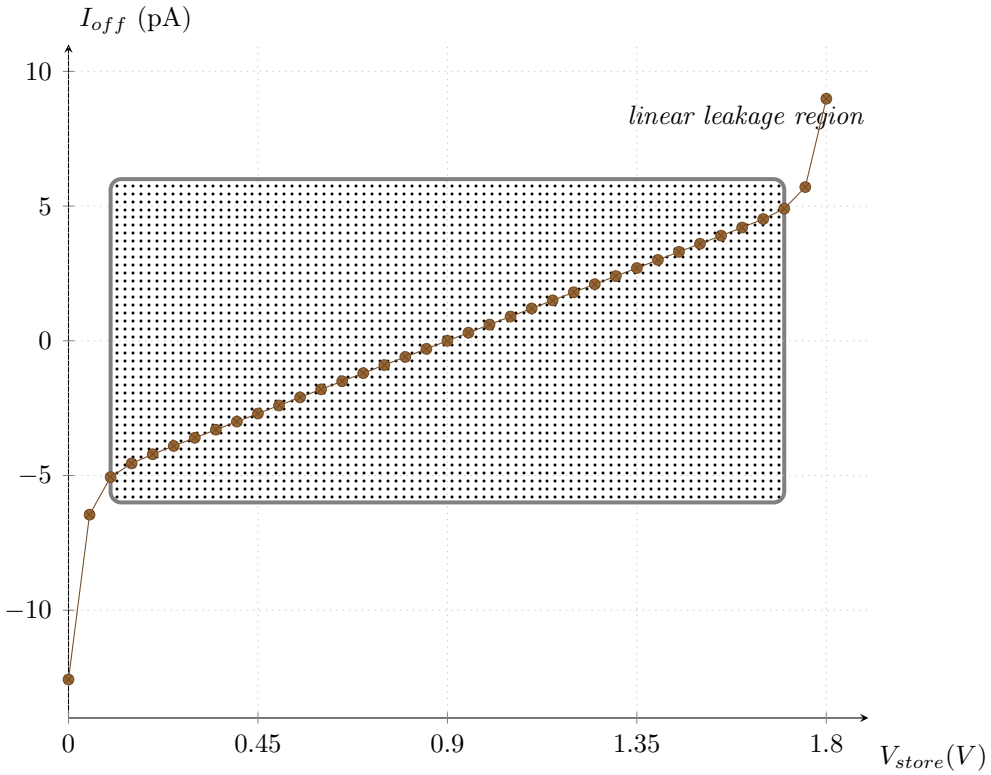


Figure 8.47 – Leakage current simulated versus the stored voltage v_S on the switch —●—, done for $v_{reset} = 0.9$.

Input decoupling The succession of two sampling switches on the input path also allows for better decoupling between the input V_{in} and the stored value V_{store}. It is illustrated in Figure 8.48. We can observe that, with this scheme, a gain of 20 dB is achieved between the simple switch architecture (Figure 8.18b) and the architecture with a *reset* (Figure 8.40).

We can observe that the input coupling is still much higher than the theoretical noise floor of the SamPic chip, thus degrading the performances of the ASIC. This simulation results can be understood and tempered with a closer look to the layout

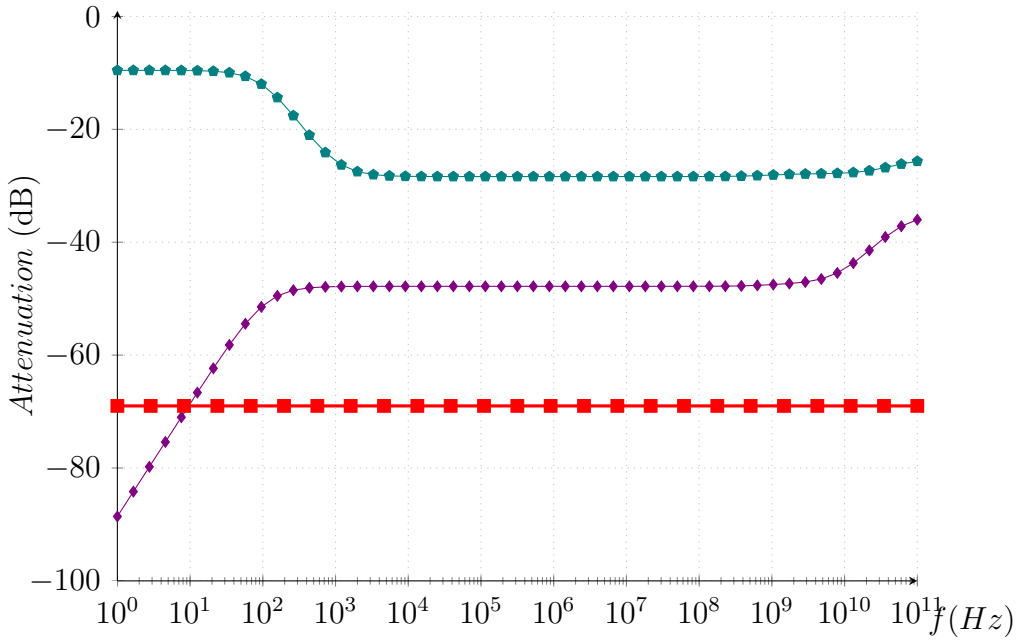


Figure 8.48 – Coupling of the input signal V_{in} to the sampling capacitor in the case of a simple switch —●— and of a sampling cell with *reset* —◆—. The kTC noise contribution of the sampling capacitor is also printed for comparison —■—.

design of the sampling cell. The layout of the switches of the sampling cells is shown in Figure 8.49.

At high frequency, the parasitic capacitors shown in Figure 8.49 become dominant and degrade the performance advantage brought by the reset switch. In order to compensate for this effect, a metal shield layer is added on top of the cell to capture any field lines going from V_{in} to V_{cap}. This shield is shown in Figure 8.50. During parasitic extraction, the algorithm used to calculate capacitance between nodes does not rely on a 3-dimensional field solver. Therefore the effect of the metal shield is not taken into account, which is the reason why a rather high coupling capacitance is observed between V_{in} and V_{cap}. In reality this capacitance is expected to be much smaller.

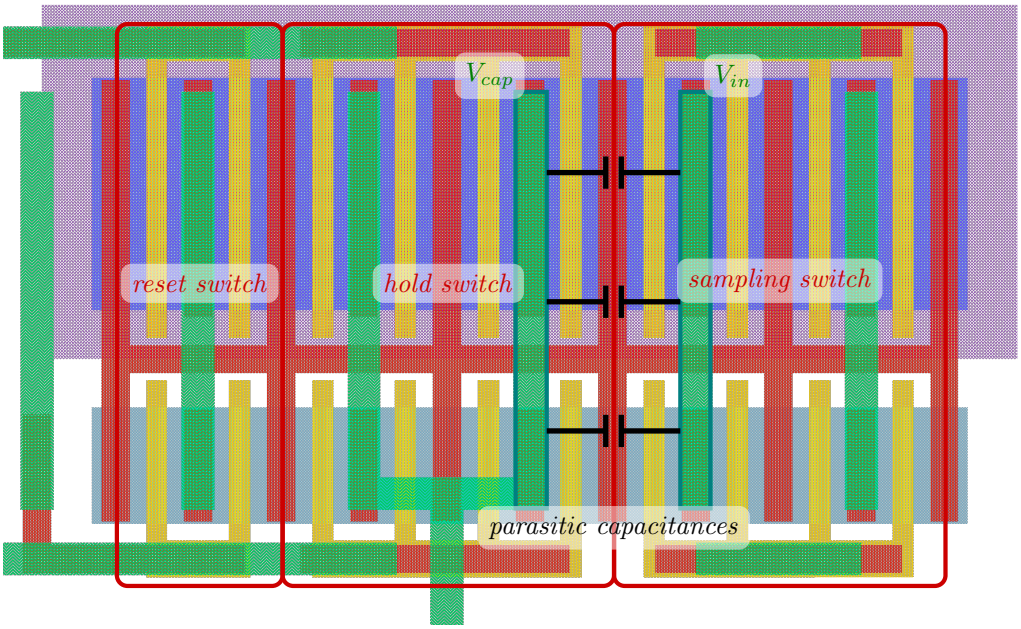


Figure 8.49 – Layout of the switches for the sampling cell. We can see that, since the metal line (in green) of the sampling switch and of the hold switch are close together, there is some parasitic capacitor between the input line V_{in} and the capacitor V_{cap}.

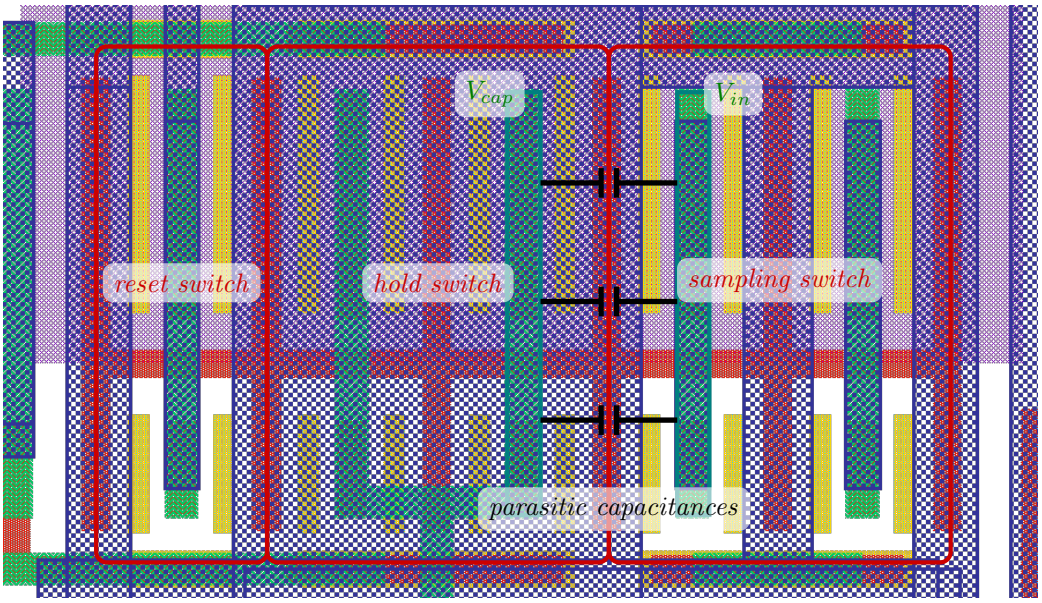
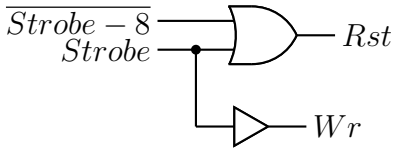


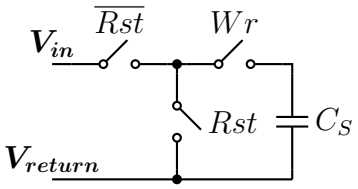
Figure 8.50 – Layout of the switches for the sampling cell with a layer of metal on top, preventing field line to go from V_{in} to V_{cap}, thus reducing the parasitic capacitance between the two nodes.

8.6.17 Sampling cell operating mode

In order to operate the sampling cell with the three steps presented in Figure 8.40, we use the control signals (*strobe*) provided by the delay locked loop (see Section 8.7 through the *trigger cell* (see Section 8.10). Only two control signals are necessary, which makes this cell simple and compact to implement although, powerful at the same time: ghost removal, input coupling reduction, leakage control.



(a) Logic used to generate the control signal of the sampling cell.



(b) Control signals used in the sampling cell.

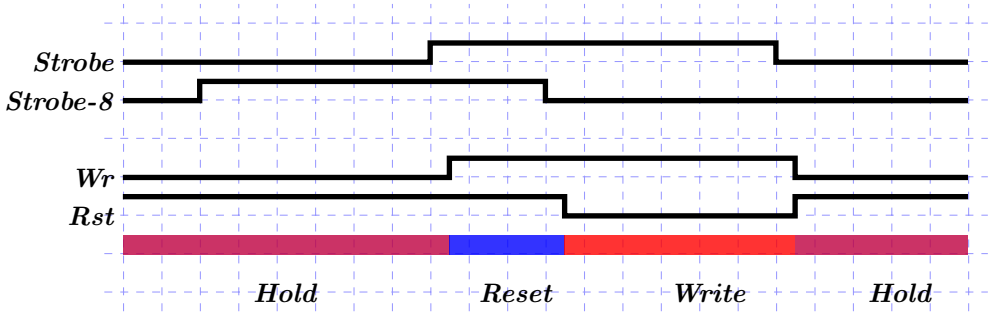


Figure 8.52 – Input and derived control signal for the sampling cell.

8.6.18 Sampling cell design

The layout of the sampling cell is shown in Figure 8.53. For clarity some layers used in the layout are not shown. It consists of the sampling switches and the capacitor. Careful shielding between the *input* line and the sampling capacitor has been done, as shown in Figure 8.50. The capacitor used is an interdigitated metal capacitor, allowing top, bottom and side shielding, in order to reduce as much as possible cross-talk effects between cells.

8.6.19 Sampling cell design summary

In summary, the sampling cell has been design following the *sampling cell with reset* architecture presented in the previous Sections. The layout of the sampling cell is shown in details in Figures 8.49, 8.50 and 8.53.

The three switches used in the cell are CMOS switches: NMOS and PMOS. I order to reduce as most as possible the resistance non-linearity of the switch (see Figure 8.27) the PMOS has been two times larger than the NMOS switch. Simuntaneously, the switch size was made substantially large in the fear of bandwith limitation of

ghosts effects. This was achieved at the cost of a higher non-linearity due to the charge injection during sampling (see Figure 8.35).

The specifications of the switch used in the SamPic cells are summarized in Table 8.6.

Table 8.6 – Summary of the sampling cell design choices.

Characteritics	Specified value
Number of switches	3
Number of control signals	2
Switch design	CMOS (PMOS & NMOS)
PMOS size	$8\mu m$
NMOS size	$4\mu m$
Bandwidth	4.8 GHz
Sampling capacitor	35 fF

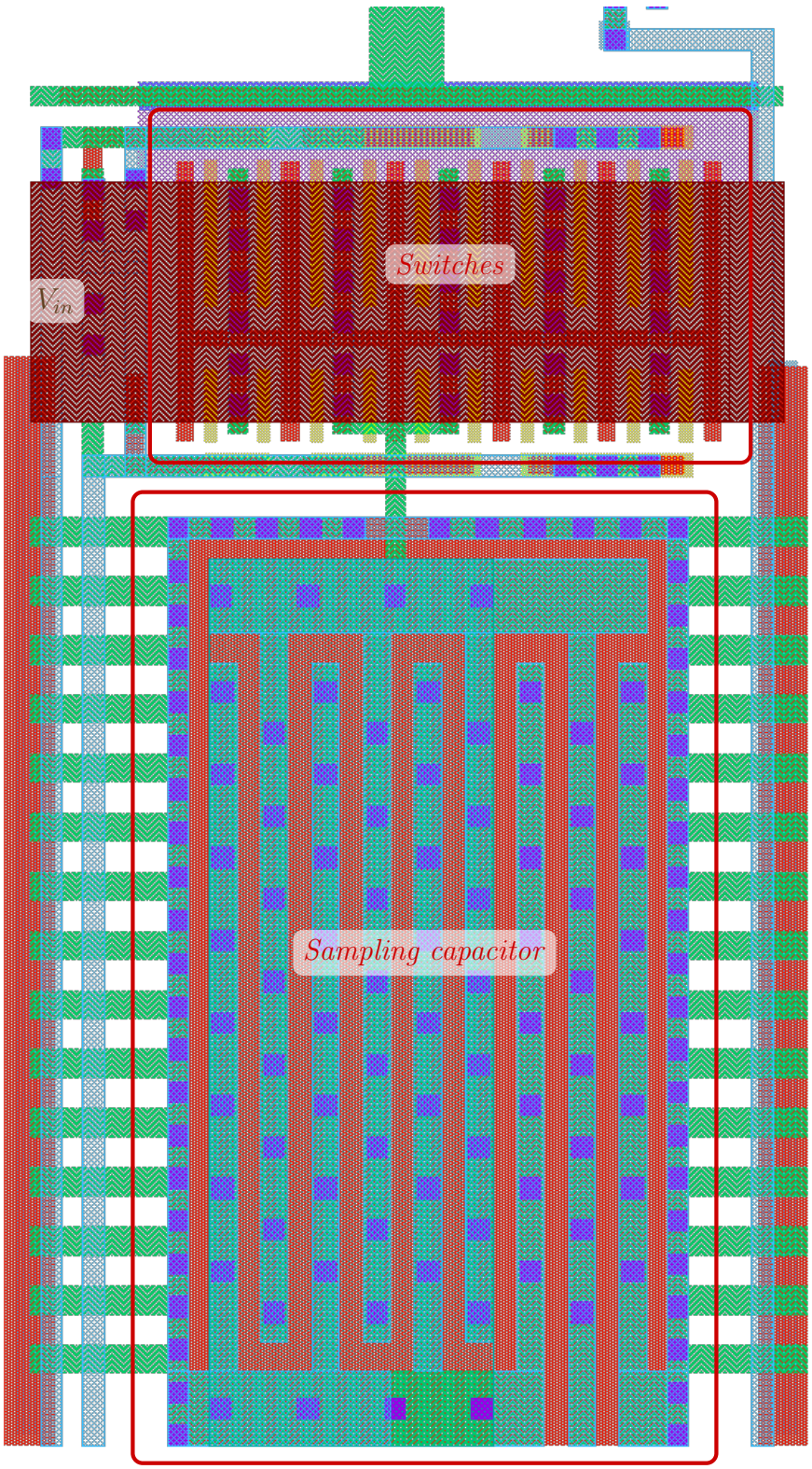


Figure 8.53 – Layout of the sampling cell, sampling switches on top and shielded sampling capacitor on the bottom.

8.7 Delay lines and delay locked loops

The delay line [6] is the structure used in the SamPic chip to create each *strobe* signal that controls the switches of each sampling cell in the array of 16 channels of SamPic. The architecture is shown in Figure 8.54. In SamPic the delay line is made of 64 cells and their output buffers.

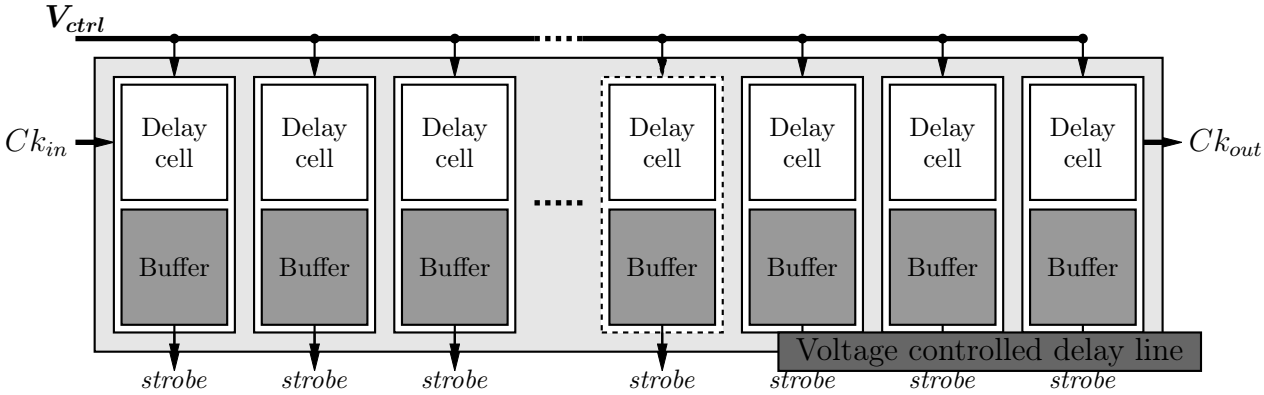


Figure 8.54 – Block diagram of the SamPic voltage control delay line. It is made of 64 delay cells producing the 64 *strobe* signals for the sampling cell arrays of the 16 channels of SamPic. In order to distribute the *strobe* signals, their outputs are buffered.

8.7.1 Delay cell

Delay cells are logical circuit elements, which delay a digital input signal by a given and adjustable value. The amount of delay can be adjusted by tuning the control voltage of the cell. This principle is illustrated in Figure 8.55.

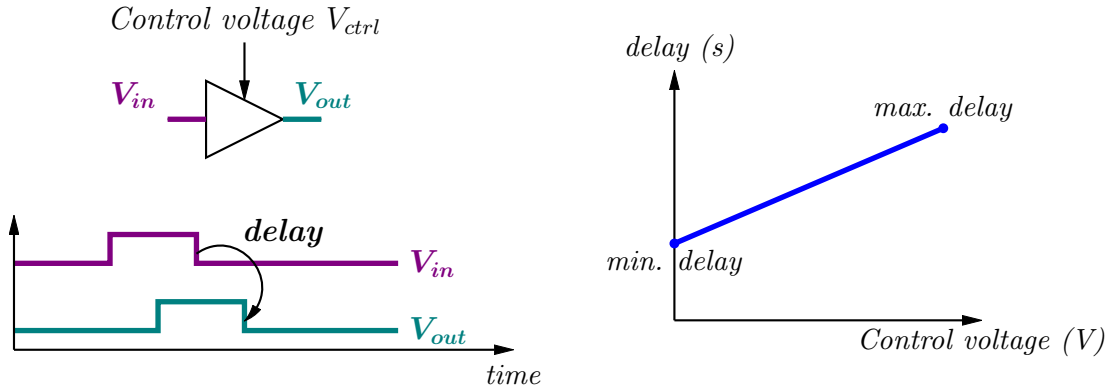


Figure 8.55 – Delay cell principle and ideal transfer curve. The cell shifts in time the input signal V_{in} by a given amount *delay*. The delay of the cell is a function of the control voltage V_{ctrl}. The range of the delay is bounded between *min. delay* and *max. delay*.

For one delay cell we have:

$$V_{out}(t) = V_{in}(t - delay) \tag{8.25}$$

For a real device, the minimum delay reach is bounded by a value, which depends on the technology used, the output load and the power consumption of the device. Smaller feature size, smaller loads and larger power consumptions allow for smaller delay.

8.7.2 Delay lines

Delay lines consist of a chain of identical adjustable delay elements, connected front to back (Figure 8.56) and controlled by the same control voltage V_{ctrl}. Since each delay element of the line is controlled by the same voltage reference, all the *delays* of the line are identical [6]. When a signal V_{in} is set at the input of a delay line made of n cells, it will propagate through the line, being successively shifted n times. At the end, the output signal V_{out} is a copy of V_{in} after $n \times delay$:

$$V_{out}(t) = V_{in}(t - n \times delay) \tag{8.26}$$

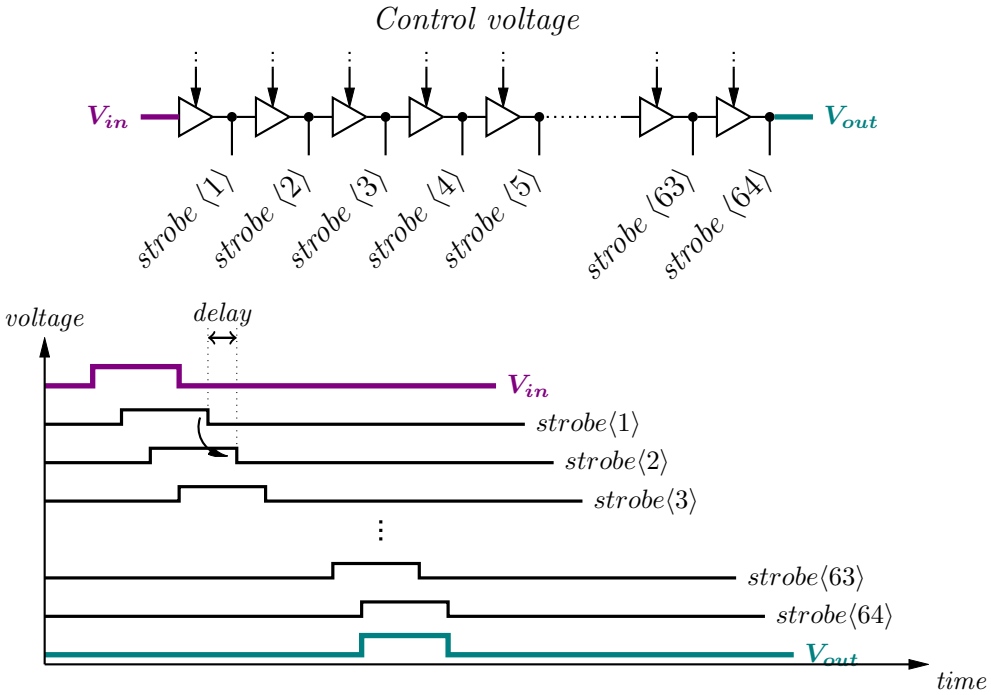


Figure 8.56 – Illustration of the working principle of a delay line. Each cell of the line shifts its input signal by a given time *delay*, which is a function of the applied control voltage V_{ctrl}. The output V_{out} comes after a time equal to $n \times delay$, since all cells delay are identical.

The principle of a delay line is illustrated in Figure 8.56. Since for an ideal delay line, all the intermediate outputs are equally shifted in time, these can be used in our *sampling capacitor array* architecture to generate the control signals: *strobes* for the sampling cells.

8.7.3 Strobe signal creation

A delay line connected to an input signal V_{in} will produce n copies of it equally spaced in time by a given time step: *delay*. So, if we assume that each strobe signal is used to sample the signal in the sampling capacitor array, the equivalent sampling frequency f will be:

$$f = \frac{1}{delay} \tag{8.27}$$

This output frequency can be tuned using the control voltage V_{ctrl} of the delay line, as shown in Figure 8.55.

This principle of simple *strobe signal creation* using a delay line and use for sampling is illustrated in Figure 8.57.

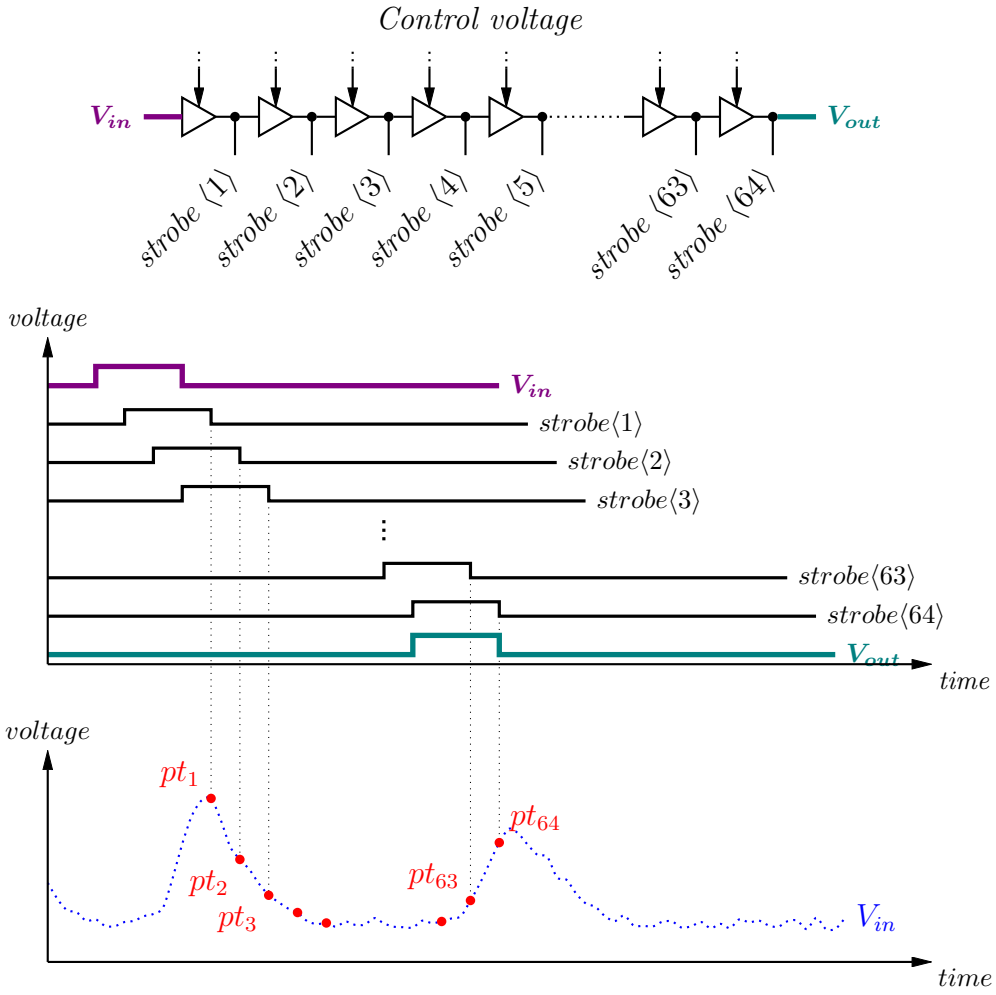


Figure 8.57 – Illustration of the use of the *strobe* signals created by the delay line to sample signals. The upper part illustrates the schematic of a voltage controlled delay line creating 64 strobe signals. The lower part illustrates the use of these strobe signals to uniformly samples 64 points (pt_1 to pt_{64}) on the input signal V_{in}.

In the CMOS technology used, as described in Section 8.3, it is possible to

design delay cells with minimum delays as low as 100 ps, corresponding to a sampling frequency f of 10 GHz and maximum delays of the order of 1 ns, corresponding to a sampling frequency of 1 GHz. So, in the end, a delay line is a simple way to control a sampling capacitor array (such as described in Section 8.6) with very high sampling frequencies: n identical *strobe* signals with sampling rates ranging from 1 to 10 GHz.

Continuous sampling If the frequency of the input signal f_{in} is chosen so that it matches the total delay of the line, we have continuous, sequential *strobe* signal creation in the delay line. In this condition, the *strobe* signals will ensure continuous, uniform sampling in time. In this case, the input signal frequency f_{in} must be set such as:

$$f_{in} = \frac{1}{n \times delay} \tag{8.28}$$

This principle is illustrated in Figure 8.58, where the input signal Ck_{in} is repeated in time with $strobe\langle 64\rangle$, so that at any given time:

$$strobe\langle 1\rangle(t) = strobe\langle 64\rangle(t - delay) \tag{8.29}$$

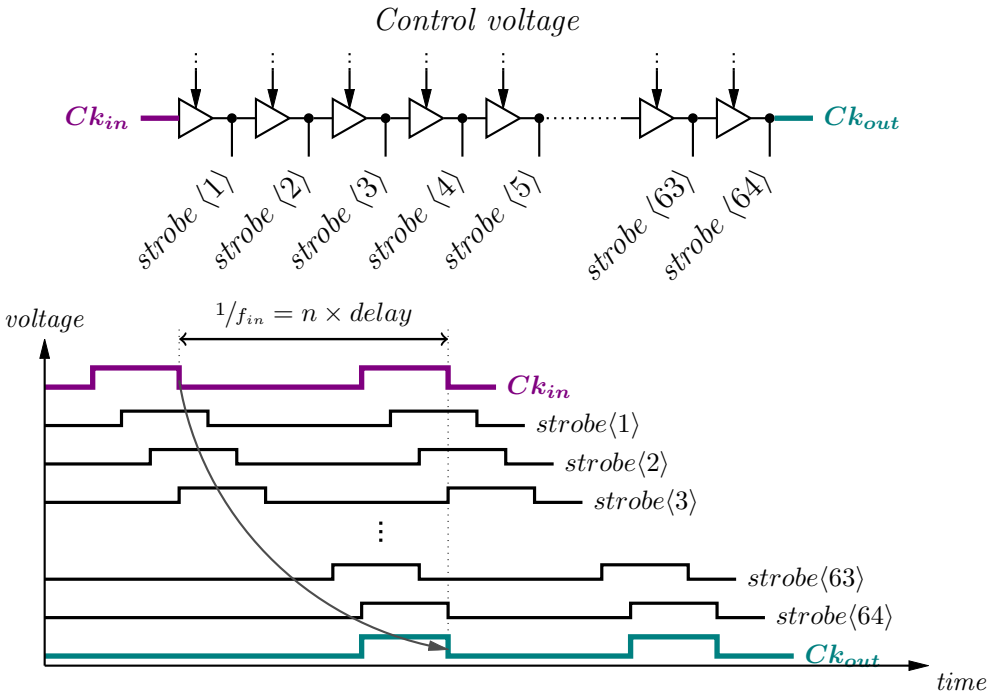


Figure 8.58 – Illustration of the delay line *strobe* signal creation for the sampling capacitor array. The input clock frequency is much less than the frequency of the 64 *strobe* signal created in parallel.

In order to verify the condition of continuous sampling written in Equation 8.28, it is necessary to control very precisely the *delay* of the cells of the delay line. This is typically done using a delay locked loop, where the total delay of the line: $n \times delay$

will be locked to a stable, external reference clock, also used as the input of the delay line.

8.7.4 Delay locked loop

A delay locked loop is a fed-back circuit where the phase of the delay line output: $n \times delay$, is servo-controlled to the phase of a reference clock Ck_{ref}. For this purpose the delay line output clock Ck_{out} is compared to the reference one Ck_{ref} using a pahse comparator delivering a phase-error signal. The phase-error is low-pass filtered to adjust the control voltage V_{ctrl} driving the delay of the line so that the phase difference stays null [6]. This is illustrated in Figure 8.59.

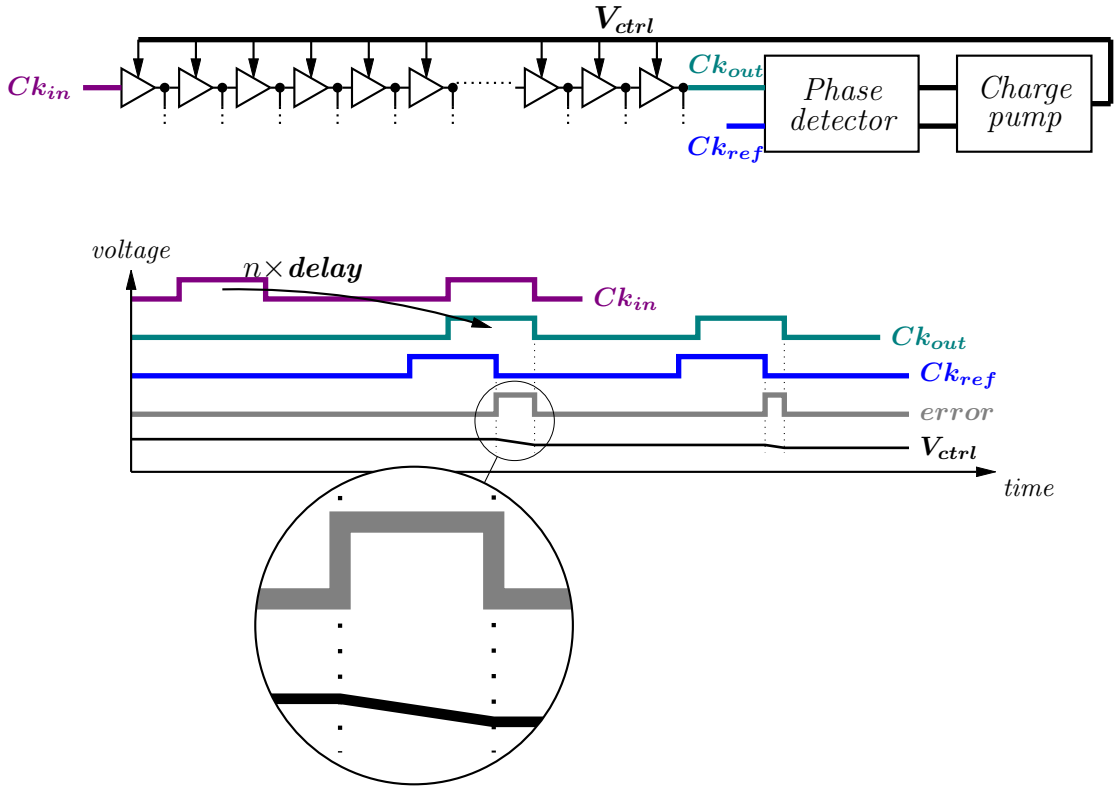


Figure 8.59 – Delay locked loop. Ck_{in} is delayed n times in the delay line. The phase detector produces a signal of width equal to the phase error between Ck_{out} and Ck_{ref}. This error signal is filtered to produce V_{ctrl} that adjusts the delay of the cells in the DLL. In this particular case Ch_{out} is coming too late, therefore V_{ctrl} is decreased to reduce the delays in the delay line.

Once locked, $\Delta\phi = 0$, meaning that the phase of the reference clock Ck_{ref} and of the output clock of the DLL Ck_{out} occurr at the same time. So we have:

$$n \times delay = {}^{1}/_{f_{ref}} \tag{8.30}$$

For example, using $n = 64$ cells servo-controlled to a 160 MHz reference clock Ck_{ref}, the sampling frequency achieved is 10.24 GSa/s. With $Ck_{ref} = 100$ MHz, the sampling frequency drops to 6.4 GSa/s.

The delay of the line is completely set by the frequency f_{ref} of the reference input clock. In particular, the variation of f_{ref}: the jitter of the reference clock will directly create some jitter on the *delay* of the line. Therefore, to get a very precise control over *delay*, typically in the order of few picoseconds, the reference clock must be very stable with jitter smaller than a picosecond. This is usually done with clock generators based on PLL followed by a jitter cleaner.

8.7.5 CMOS Delay cell architecture

Current starved inverters In CMOS technologies, a delay cell can be made of a chain of two inverters. However the *delay* of a basic CMOS inverter cannot easily be controlled. An adjustable delay is then realized by starving the supply current of the inverters [6]. This is done by adding transistors in the power supply path. This is illustrated in Figure 8.60, where the transistors T_N and T_P, controlled by $Ctrl_N$ and $Ctrl_P$, are added.

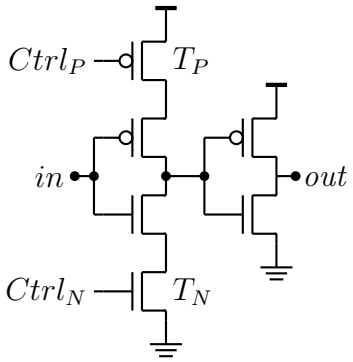


Figure 8.60 – Schematic of a current starved delay element with control on the rising $Ctrl_N$ and falling $Ctrl_P$ edge of the input signal in.

Current limitation The maximum current I_D that can flow through an NMOS transistor in saturation region, versus the *gate-source* voltage v_{GS} applied, is shown in Figure 8.61. Above V_T, the saturation current I_D is a quadratic function of the *gate-source* voltage v_{GS}. Therefore, in this region, the transistor can be seen as a voltage controlled current limiter.

For a NMOS transistor [6]:

$$I_D = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{GS} - V_T)^2 \tag{8.31}$$

Delay control The operation principle of controllable delays using current starved inverters is illustrated in Figure 8.62. Considering a signal going from high to low at the input of the cell (Figure 8.62a), the current flows from vdd to the node I via T_P. The charging time of the node I is controlled by the maximum amount of current allowed through T_P, via $Ctrl_P$. After the *charging time* the edge is inverted by the

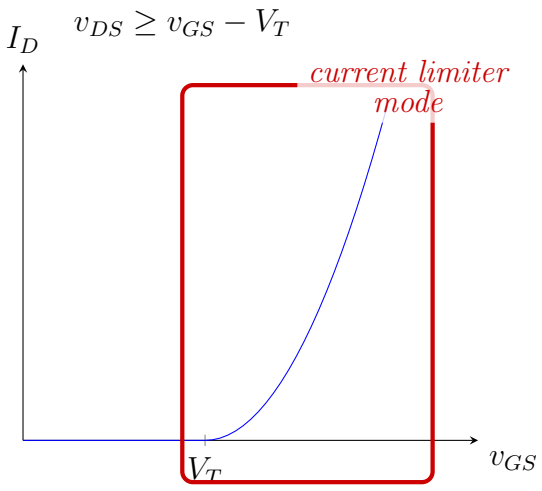


Figure 8.61 – Transconductance characteristic of the T_N transistor.

second inverter of the delay cell, introducing a *falling delay* between the input and the output edges.

Similarly $Ctrl_N$ controls the delay of the rising edge in the delay cell.

8.7.6 Noise considerations

Control of the charge and discharge currents, and hence the delay of the current starved inverter is directly set by v_{GS} on the transistors T_N and T_P. A noisy v_{GS} will deliver a non-uniform current I_D to the internal node I of the delay cell, and thereby a non-uniform charging and discharging times from cycle to cycle, which will create jitter on the *rising* and *falling* edges of the delay cell output. This effect is sketched in Figure 8.63.

Developing the expression of v_{GS} for T_N we have:

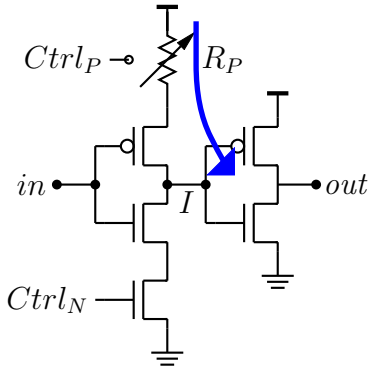
$$v_{GS} = Ctrl_N - gnd \tag{8.32}$$

where, for T_P:

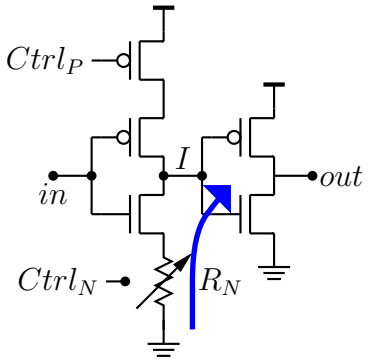
$$v_{GS} = vdd - Ctrl_P \tag{8.33}$$

In CMOS technology *gnd* can easily pick-up noise by substrate coupling, whereas *vdd* can be made very clean using a proper isolation and a dedicated chip input. Therefore we avoid using NMOS starved inverters. Additionally it is more complicated to make a delay locked loop where the delays of both rising and falling edges of the cells must be controlled.

Based on these considerations, a new type of delay cell has been developed and is presented. A delay cell without an NMOS current limiter is shown in Figure 8.64.



(a) Current paths during a falling edge.



(b) Current paths during rising edge.

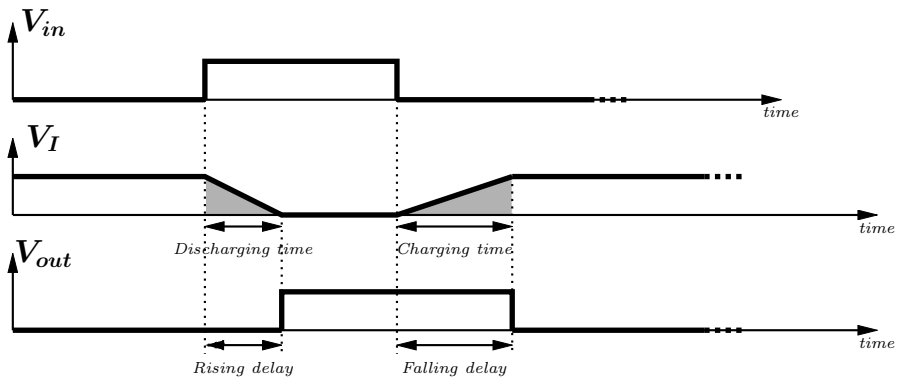


Figure 8.62 – Simplified illustration of the delay cell operation principle. V_{ctrlP} operates on falling edges, limiting current flowing downwards, and increasing the time required to charge the node I (Fig 8.62a), while V_{ctrlN} operates on rising edges by limiting upwards currents and increasing the time required to discharge the node I (Fig 8.62b). Reducing the current flow increases the charging time of the internal node thus the delay of the cell.

8.7.7 Single edge delay cell

A delay cell with a single current limiter is shown in Figure 8.64. As described previously, it can only control the delay of *falling edges*, with the voltage control $Ctrl_P$. The *rising edge delay* does not change with $Ctrl_P$.

From the formula of the saturation current I_D in the P transistor [6], we can derive a formula for the controllable resistance R_P of the switch versus the control voltage $Ctrl_P$:

$$R_P = \frac{1}{\mu_p C_{ox} \frac{W}{L}(vdd - Ctrl_P - V_T)} \tag{8.34}$$

R_P is a highly non-linear function of $Ctrl_P$ as it varies proportional to $^1/_{Ctrl_P}$. As described in Figure 8.62a, the charging time of node I during a *falling edge* controls the *falling edge delay* of the cell. The charging time is controlled by the RC filter made by the resistance R_P of the switch and C_I, the capacitor of node I. This is illustrated in Figure 8.65.

Therefore, during an input falling edge, the voltage $V_I(t)$ on node I, rises from

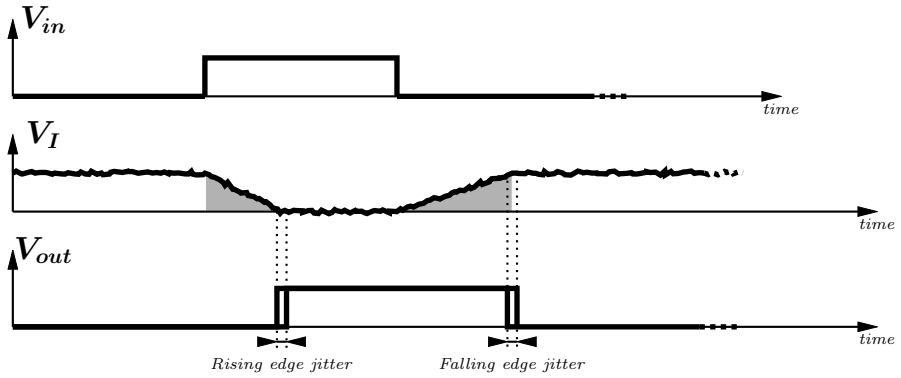


Figure 8.63 – Illustration of the creation of jitter in the delay cell due to noise on the v_{GS} control voltage of the current starved inverters. Noise on v_{GS} couples to the cell internal node I, through the transconductance of transistors T_N and T_P.

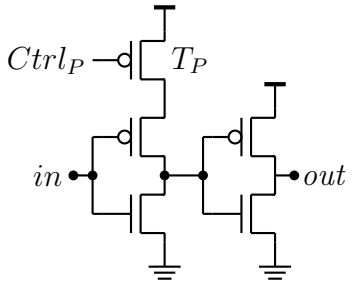


Figure 8.64 – Current starved delay element with control only on the falling edge of the input signal.

gnd to *vdd* and is given by:

$$V_I(t) = vdd\left(1 - e^{\frac{-t}{R_P C_I}}\right) \tag{8.35}$$

We can get an idea of the *delay* introduced by the transistor T_P by calculating the time when V_I reaches $^{vdd}/_2$.

$$\begin{aligned} delay &= 2 \times R_P C_I \\ &= \frac{2C_I}{\mu_p C_{ox} \frac{W}{L}(vdd - Ctrl_P - V_T)} \end{aligned} \tag{8.36}$$

As expected, the *delay* introduced by the current-starved inverter is highly non-linear. The simulation results of the *rising edge* and *falling edge* delays versus the control voltage $Ctrl_P$ of the single-edge delay cell are reported in Figure 8.66. The rising edge delay does not depend on $Ctrl_P$.

8.7.7.1 Operating mode

If we consider a chain of delay cells with falling edge delay control and with a rising edge delay so that: $delay_{fall} \geq delay_{rise}$, two cases can be identified. They are illustrated in Figure 8.67.

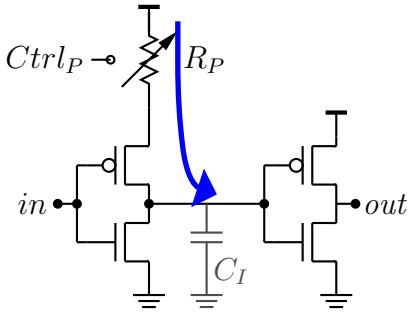


Figure 8.65 – Illustration of the RC stage made of R_P, the PMOS resistance and C_I the capacitor of node I. During a *falling edge* C_I is charged up trough R_P.

Positive input pulse An input pulse rising then falling is sent at the input of the delay line. The rising edge delay, being smaller than the falling edge delay, the pulse width *increases*, the effect is described in Figure 8.67a. After n cells, if $width_{in}$ was the width of the input signal, the width $width_{out}$ of the output signal is given by:

$$width_{out} = width_{in} + \sum^{n} delay_{fall} - delay_{rise} \tag{8.37}$$

Negative input pulse An input pulse falling the rising is sent at the input of the delay line. The falling edge delay being larger than the rising edge, the pulse width *decreases*, the effect is described in Figure 8.67b. After n cells, when $\sum^{n} delay_{rise} - delay_{fall}$ becomes smaller than $width_{in}$ and the pulse disappears.

So, in order to always have pulses travelling the delay line, we must always be in the situation where for any control voltage on the delay cells:

- For positive input pulse: $delay_{fall} \geq delay_{rise}$
- For negative input pulse: $delay_{rise} \geq delay_{fall}$

During normal operating mode, single edge delay cells have the disadvantage of having traveling pulses with increasing *width*, as illustrated in Figure 8.67a. In Figure 8.12 of Section 8.6, we saw that the sampling cell is in *tracking mode* when the control signals from the delay line are high. Non-uniform output *widths* from the delay cell yield to non-uniform *tracking* time for the sampling cells non-uniform *residual charges* (see Figure 8.36).

In SamPic, we have designed a delay cell based on the logical NOR gate, and the single edge delay cell produces constant width outputs.

8.7.8 Nor based delay cell

The cell of Figure 8.64 has the disadvantage of making pulses with width growing in a delay line (Figure 8.67a), and a highly non-linear delay response function of its control voltage (Figure 8.66) [7]. An alternative to avoid this issue is to use the delay cell architecture of Figure 8.68. This cell brings two improvements comprared to the one previously described:

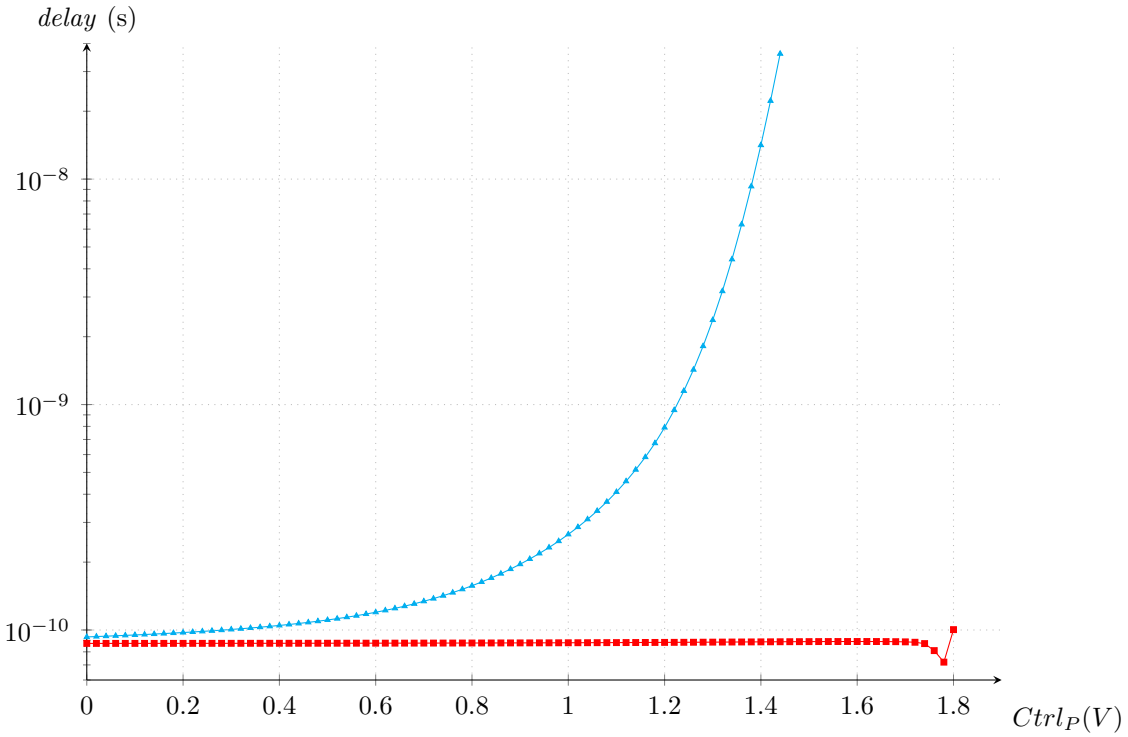


Figure 8.66 – Simulation of the *rising edge* delay and *falling edge* delay of the single-edge delay cell versus the control voltage $Ctrl_P$. The rising edge delay does not vary with $Ctrl_P$, while the falling edge delay does, with highly non-linear dependence.

8.7.8.1 Fixed size sampling window

The current starved P-transistors are now placed on the second inverter of the delay cell. Therefore, the control voltage now delays *rising edges*.

Since we are now delaying rising edges with this cell, negatives pulses must be fed to the delay line to avoid pulse disappearance.

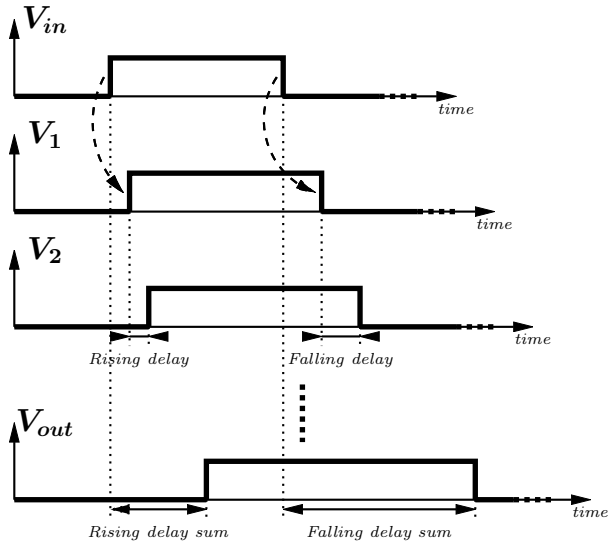
Operating mode The input stage instead of being a simple inverter is now replaced by a CMOS *NOR* gate. The timing diagram for operation of this cell is illustrated in Figure 8.69.

At the input of the *NOR* gate we have V_{in} and $\overline{V_{in-12}}$. $\overline{V_{in-12}}$ is the inverted value of the 12^{th} previous cell in the delay line. If *delay* is the amount each cell of the delay line shifts its input signal, we have:

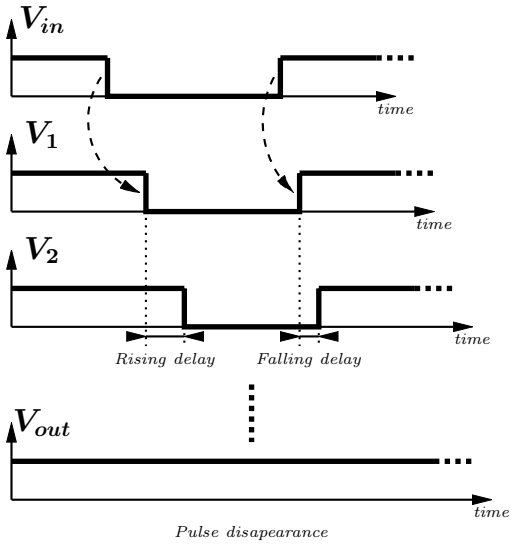
$$\overline{V_{in-12}}(t) = V_{in}(t - 12 \times delay) \tag{8.38}$$

When both V_{in} and $\overline{V_{in-12}}$ are equal to 0 a positive signal is created at node I. When V_{in} goes back to 1, V_I returns to 0.

Subsequently this pulse is propagated through the current starved inverter, which delays more the falling edge than the rising.



(a) Positive pulse propagation in the chain. Case where $delay_{fall} \geq delay_{rise}$.



(b) Negative pulse propagation in the chain, similar case. The pulse disappears completely after a few cells.

Figure 8.67 – Pulse propagation in a chain of single edge (falling) controlled element

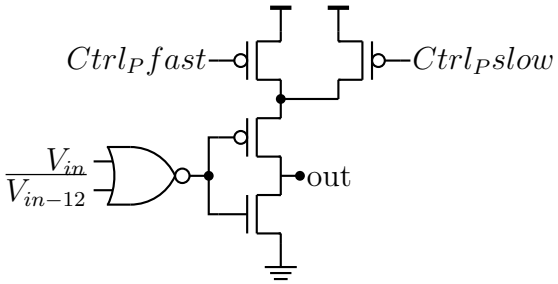


Figure 8.68 – Current starved delay element with fixed window and two different PMOS for current starving.

Constant width If we consider that the control voltage $Ctrl_P$ of the cell gives: $delay_{rise} = delay$ and that we have $delay \geq delay_{fall}$, we prove by induction that we have a constant width pulse in the delay line.

We are controlling the rising edge delay, so by definition the **rising edge of** V_{in} **comes exactly** $12 \times delay$ **after** V_{in-12}.

Also because we have: $delay \geq delay_{fall}$, the falling edge of V_{in} comes slightly faster than its rising edge, so the *width* of V_{in} is larger than $12 \times delay$, let's make it:

$$width\,(V_{in}) = 12 \times delay + cste$$

In the NOR cell, the rising edge of V_{in-12} and the rising edge of V_{in} are defining the new window. Since the rising edge of V_{in} comes exactly $12 \times delay$ after V_{in-12}, we have:

$$width\,(V_{in} \vee V_{in-12}) = 12 \times delay$$

The output of the *NOR* gate is then fed to the current starved inverter, which increases the width of the pulse by $delay - delay_{fall}$ so:

$$width\,(V_{out}) = 12 \times delay + delay - delay_{fall}$$

Therefore, the width of the output pulse in the delay line is a constant equal to:

$$12 \times delay + cste \text{ (with: } cste = delay - delay_{fall})$$

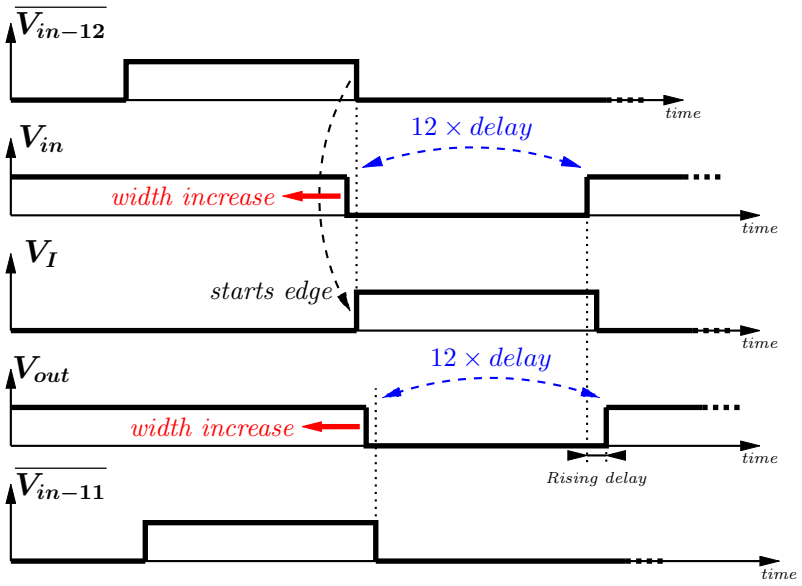


Figure 8.69 – Working principle of the NOR delay cell of Fig. 8.68. The cell input consists of V_{in} and $\overline{V_{in-12}}$. By construction $delay_{rise} \geq delay_{fall}$, so the width of the input edge increases on his falling edge side. The rising edge of V_{in-12} starts the edge of the window on the node I. The propagation of the rising edge is controlled by V_{Ctrl}.

Advantages The advantages of such structure are multiple:

- Identical pulses propagating at each time step in the DLL: uniform power consumption in the line.
- Pulses with constant width: a constant number of cells active during sampling.
- Pulses with constant width: constant charging time in the sampling cells.
- Pulses generated with V_{in-12}: possibility to create a delay line loop with no dead zones.
- DLL sampling frequency controlled by single input voltage.

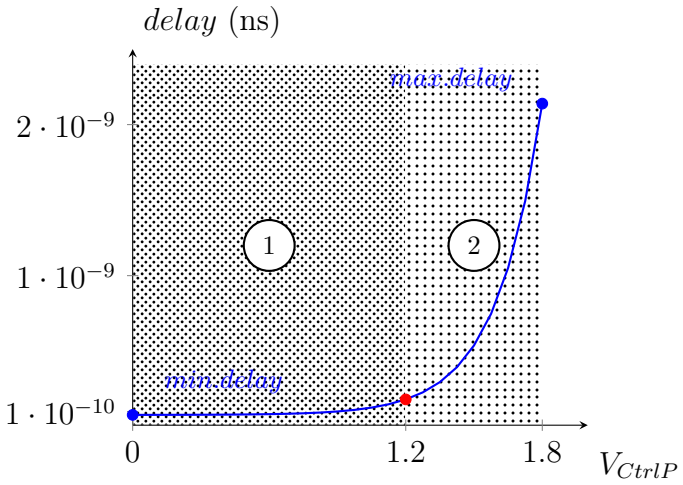


Figure 8.70 – Plot identifying the domain separation that can be done. Domain 1: region where the delay varies slowly with the control voltage. Domain 2: region where the delay varies quickly with the control voltage.

8.7.8.2 Linearization of the delay control

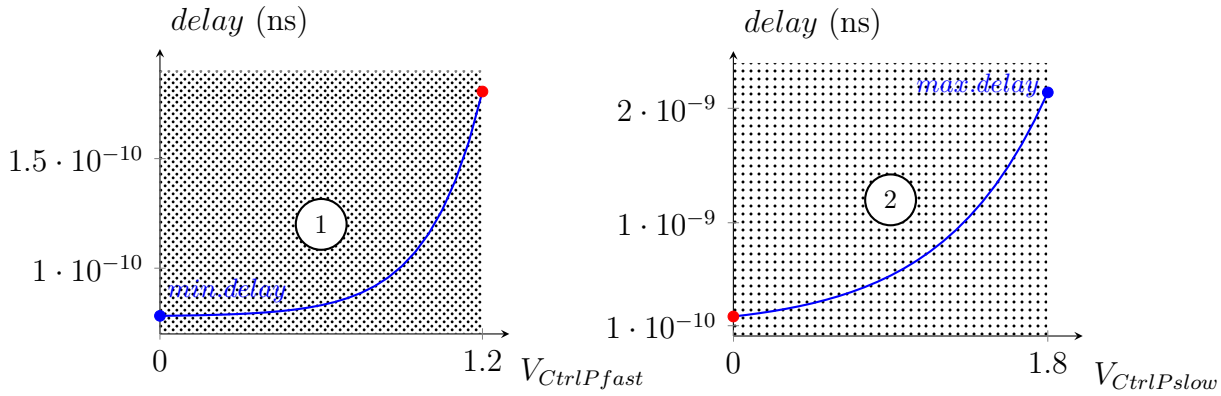
The delay of the cell function of the control voltage is highly non-linear (see Figure 8.66). A fine control of the delay over the whole range of $Ctrl_P$ is difficult, we can roughly identify two regions in Figure 8.71a.

- *Region 1* corresponds to a domain where the resistivity of the PMOS transistor is low, a lot of current is allowed to the inverting branch of the delay cell.

- *Region 2* correspond to a domain where the resistivity of the PMOS rises exponentially, and a small variation of the control voltage will produce a large variation in the cell delay.

This situation is not ideal to control precisely the delay line, especially in *Region 2*, where a small perturbation on the control voltage will yield to large variations on the cell delay. To improve this situation an additional Pmos starving transistor is inserted in parallel of the existing one as shown in Figure 8.68. This additional control transistor is made more resistive than the first one to target directly *Region 2*. This is illustrated in Figure 8.71, the two control regions of Figure 8.70 are divided in two by the use of the two transistors controlled separately by $Ctrl_P fast$ and $Ctrl_P slow$. When using $Ctrl_P fast$, the small delay region can be reached with low control voltages. At higher values of $Ctrl_P fast$, the delay cell enters *Region 2* and the control voltage is switched to $Ctrl_P slow$ which gives a more accurate control.

The advantage of this linearization technique is to be able to reach very fast (10 GSa/s) and much slower (500 MSa/s) sampling speeds with an equally good sampling precision (picosecond precision).



(a) Plot of Region 1: plot of the region where the use of $V_{CtrlPfast}$ allows for precise control of small delays.

(b) Plot of Region 2: delay function of the control voltage $V_{CtrlPslow}$ when using additional more resistive transistor for current starving.

Figure 8.71 – Zoom of the two region of Figure 8.71a

8.7.8.3 Delay line architecture

The SamPic chip delay line is made of 64 delay cells in order to control the 64 cells of the sampling capacitor array. Since the NOR based delay cell uses V_{in-12}, 12 dummy delay cells are needed before the 64 cells. 12 dummy cells are also added at the end of the line to ensure identical capacitive loading for all 64 cells of the delay line. Part of the delay line layout is shown in Figure 8.72. We saw in Figure 8.68 that the capacitive load of node I is crucial to control the *delay* of the line. To get precise timing control out of the delay line it is therefore essential to control extremely precisely all capacitances in the line. Therefore, the layout of the cell is made extremely reproducible from cell to cell.

8.7.9 Delay line design summary

Since the delay cells constituting the delay line are the key to the sampling precision in SamPic sampling capacitor arrays, a carefull attention was brought to the delay line design in order to maximize the stability and precision of its delay eleements.

The delay line is designed to deliver strobe signal with frequencies ranging from 1 GSa/s to 10 GSa/s to the sampling cells. The locking to these sampling frequencies is achieved internally using a phase comparator and a charge pump locked to an external reference clock.

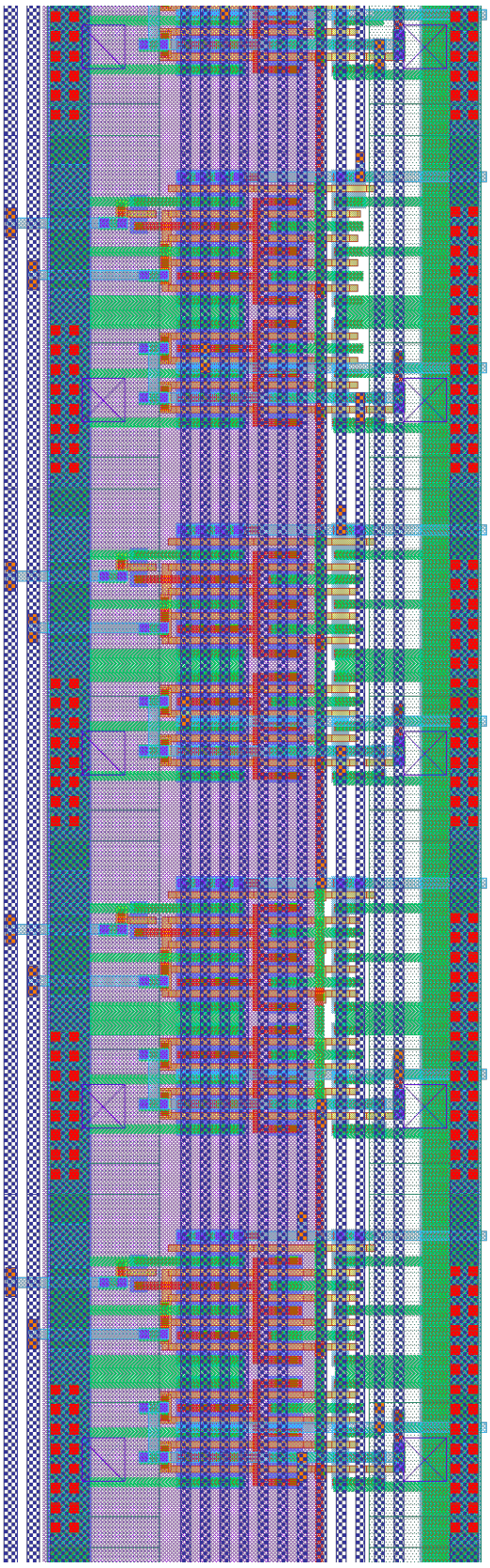


Figure 8.72 – Part of the layout of the DLL

8.8 Phase error correction

As illustrated in Figure 8.59, in order to get precise control over the sampling frequency, the total time shift of the delay line: $n \times delay$ must be locked to an external reference clock.

8.8.1 Correction principle

Correction of the phase error between the *output* signal of the delay line and the *reference* signal is performed using a phase comparator, followed by a charge pump as illustrated in Figure 8.59 [6]. The phase comparator detects the phase difference between the input signal and the reference and the charge pump adjusts the control voltage of the delay line to compensate for this phase error. The awaited result of the correction is illustrated in Figure 8.73.

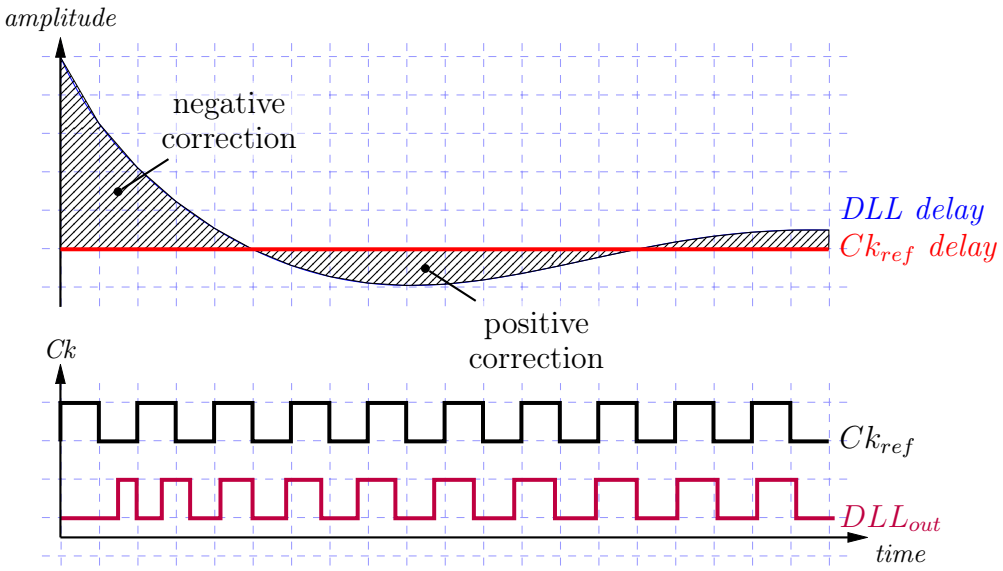


Figure 8.73 – Principle of the phase detection correction

Due to the time required for the servo-controlled phase error correction to become effective (*i.e.* for the phase error to become negligible), there is a delay before sampling can be started inside SamPic. This delay is present at each delay line startup.

8.8.2 Basic principle of the charge pump

A charge pump is a very simple circuit, which adjusts the control voltage on the delay line. The schematic of the charge pump circuit used in SamPic is shown in Figure 8.74. Small voltage pulses, *Up* and *Down* are sent on the transistors, charging and discharging the C_P capacitor, thus raising or lowering the control voltage $Ctrl_P$.

This operating principle is illustrated in Figure 8.75, where the *Up* or *Down* voltage pulses are increasing or decreasing $Ctrl_P$. When there are no pulses on *Up* nor *Down*, the control voltage remains constant. This means that when the

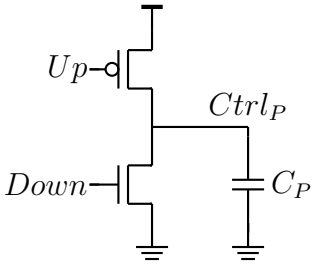


Figure 8.74 – Illustration of the structure of the charge pump. The voltage $Ctrl_P$ of the delay line is driven by current injection of the capacitor C_P via the two transistors.

control voltage has reached the value where Equation 8.30 is verified, and no voltage correction on $Ctrl_P$ are necessary, C_P holds a very clean, stable control voltage.

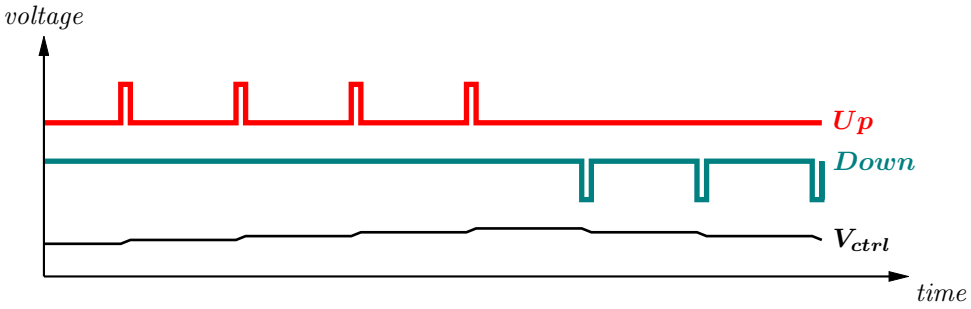


Figure 8.75 – Illustration of the operating principle of the charge pump. Up pulses are increasing $Ctrl_P$, $Down$ pulses are decreasing $Ctrl_P$.

The Up and $Down$ control voltage are driven by the phase comparator to achieve locking of the delay line at the proper delay. They act as a low pass filter.

8.8.3 Basic principle of the phase detector

8.8.3.1 Correcting the phase error

The objective of the phase detector is to detect phase differences between the input signals and deliver the correction signals Up and $Down$ to the charge pump. The illustration of the phase detector effect on the delay of the delay line *DLL delay*, as shown in Figure 8.73. At startup, the *DLL delay* is far from the value of the reference clock period Ck_{ref} *delay*, by successive corrections the delay of the line is brought equal to the reference clock period [6].

In order to bring the phase difference between its two input signals to zero, the phase comparator commands a *delay increase* when the phase difference is negative (by pulsing on the Up control signal). Similarly, when the phase difference is positive, the phase comparator commands a *delay decrease* (by pulsing on the $Down$ control signal). In order to improve the convergence time of the processus, the amount of correction is made larger for phase differences far from the zero-phase stability point. This principle is illustrated in Figure 8.76, where we have larger corrections far away from the stability point *phase delay* = *0*.

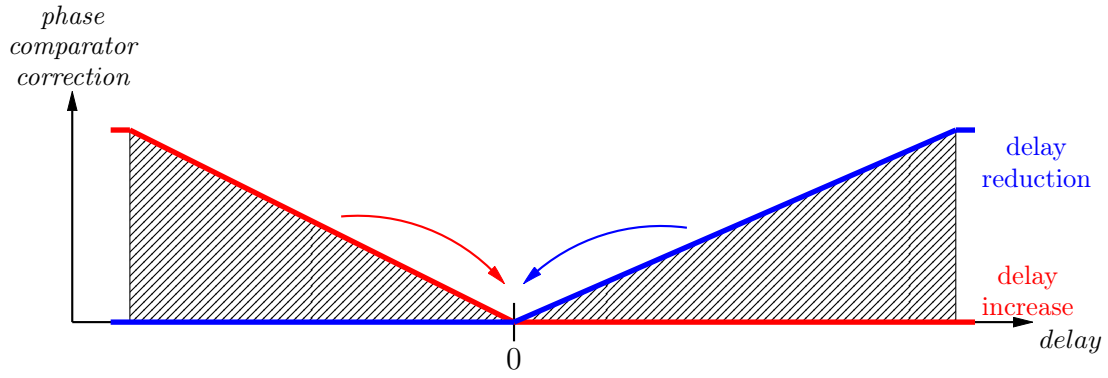


Figure 8.76 – Principle of the phase detector

A simple phase comparator can be made of an RS NAND latch. The input pulse of the DLL is fed to the $\bar{S}$ input of the latch and the last pulse out of the DLL to the $\bar{R}$ input. The outputs of the latch are fed to the charge pump (see Figure 8.74)

To adjust the inevitable deadtime between the arrival of the reference clock or the input clock, the user can choose to delay one or the other.

The correction of the detector is proportional to the time difference between *DLL delay* and Ck_{ref} *delay*, with a negative correction (*i.e. DLL delay* decrease) when:

$$DLLdelay < Ck_{ref}delay \tag{8.39}$$

and a positive correction (*i.e. DLL delay* increase) when:

$$DLLdelay > Ck_{ref}delay \tag{8.40}$$

In order to control precisely the sampling frequency of the delay line, the total delay of the delay line needs to be controlled so that $64 \times delaytime = refclock$. For a sampling frequency of 10 GSa/s, $delaytime = 100$ ps and the reference clock period needs to be equal to the total delay time of the delay line: 6.4 ns. In order to do so, an internal phase comparator compares the output phase of the delay line to the one of the reference clock. In order to minimize the jitter, locking time and dead time of the phase comparator, the comparison must be done as fast as possible.

8.8.3.2 Phase comparator dead zone

Theoretically, the response of an ideal phase comparator follows the principle shown in Figure 8.76. In reality, when the phase difference of the two inputs: DLL_{out} and Ck_{ref} starts to be really small, propagation delays in the circuit create a *dead zone*, where a comparison is no more possible. These delays are typically in the order of few hundreds of picoseconds, corresponding to the gate transmission times. This is not acceptable for a circuit where we are trying to measure times at the picosecond level.

The response of a phase comparator with dead zone is illustrated in Figure 8.77. Within the dead zone it is not possible to adjust the delay line sampling frequency, creating jitter in the structure.

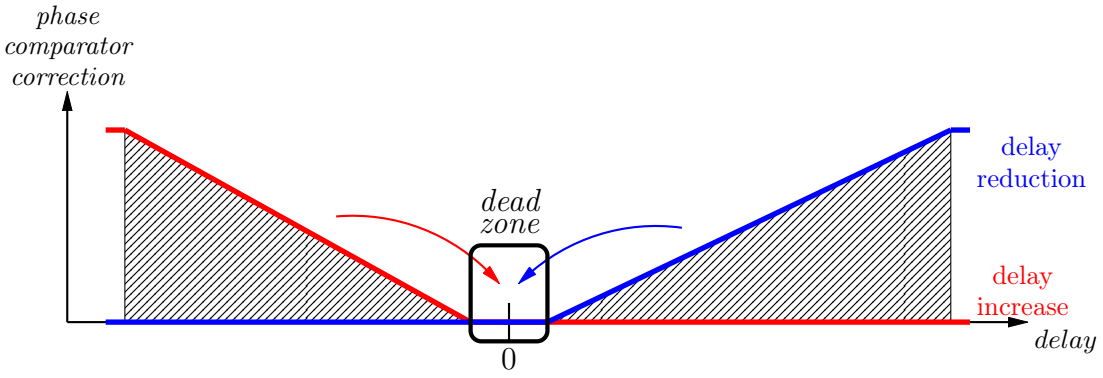


Figure 8.77 – Illustration of the dead zone in the *RS* phase comparator.

8.8.4 Asymmetrical phase comparator

A simple solution to the *dead zone* issue is to make the control signal of the phase detector asymmetrical. This is illustrated in Figure 8.77. Here the control signal *Down*, used for delay reduction in the delay line is made constant, always on no matter what the phase delay is. The *Up* signal stays as previous. The dead zone of the phase comparator is effectively suppressed, however at the cost of a small displacement of the lock-in frequency. The principle of a phase comparator with asymmetrical delay correction is illustrated in Figure 8.78.

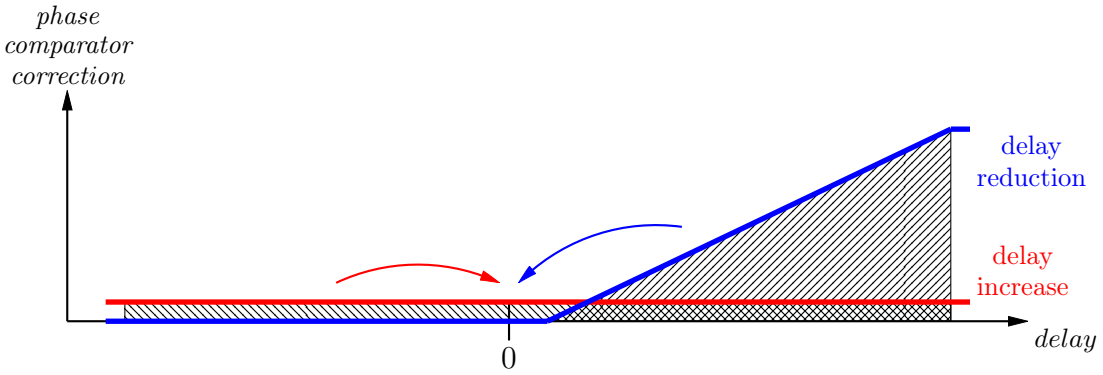


Figure 8.78 – Principle of the asymmetrical phase detector, the amount of correction given by the phase comparator is a fixed value for negative phase difference in this case.

8.8.4.1 Principle

In the asymmetrical architecture used in SamPic only the *Up* pulse is created based on the phase difference between the input of the delay line DLL_{in} and the output DLL_{out}. When DLL_{out} is in advance *Up* is set to *one* and reset (with priority) by DLL_{in}. This can be achieved using a *RS* NAND latch which has a priority reset. The process to build the *Up* and *Down* signals is illustrated in Figure 8.79.

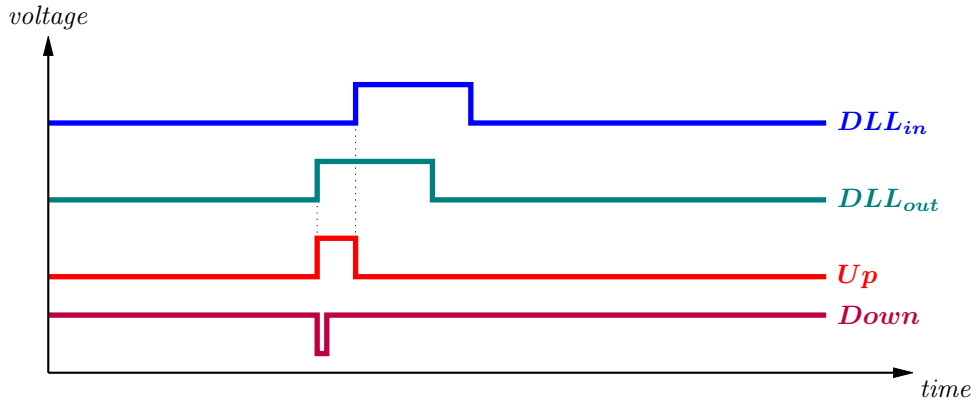


Figure 8.79 – Illustration of the *Up* and *Down* pulse formation in the SamPic phase comparator. When DLL_{in} is in advance, *Up* is set to *one* and reset with priority by DLL_{out}. At the same time the *Down* pulse is constantly created.

8.8.4.2 Architecture

The architecture of the asymmetrical phase comparator is shown in Figure 8.80. For negative phase differences, the amount of correction is always the same, created by a pulse maker on the delay line output, whereas the correction for the positive delay is done by the *RS* latch.

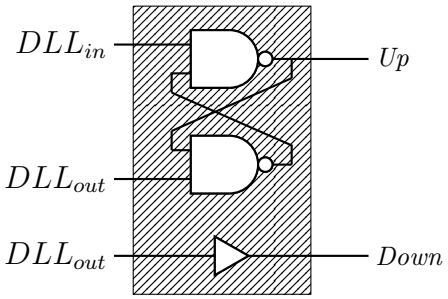


Figure 8.80 – Architecture of the asymmetrical phase detector. For negative phase differences, the amount of correction is always the same, created by a pulse maker on the delay line output at each falling edges, whereas the correction for the positive delay is done by the *RS* latch.

8.8.5 Conclusion

The phase comparator presented in this Section and designed in SamPic presents the advantage of simplicity, both of operation and design along with great performances: absence of dead zone, low jitter, which makes it a great candidate for our picosecond sampling ASIC.

8.9 ADC conversion

8.9.1 Introduction

The role of the Analog to Digital Converter (ADC) inside SamPic is to convert the analog sampled values into a digital representation. During the analog-to-digital conversion of a SamPic channel, the sampling process must be stopped in that channel, making it inoperative (or "dead"). Therefore, the conversion must be performed as fast as possible to minimize the channel "dead-time" due to the conversion time.

8.9.2 SamPic Wilkinson ADC

Conversion of the sampling cell analog value is done by an 11-bit Wilkinson ADC [8], [5], one per sampling cell. Their structure inside the channel is shown in Figure 8.81, they are made of one comparator and 11 identical digital memory cells storing 1bit. The advantage of the Wilkinson ADC is to perform the conversion in parallel for all the cells of one channel, thus reducing the "dead-time".

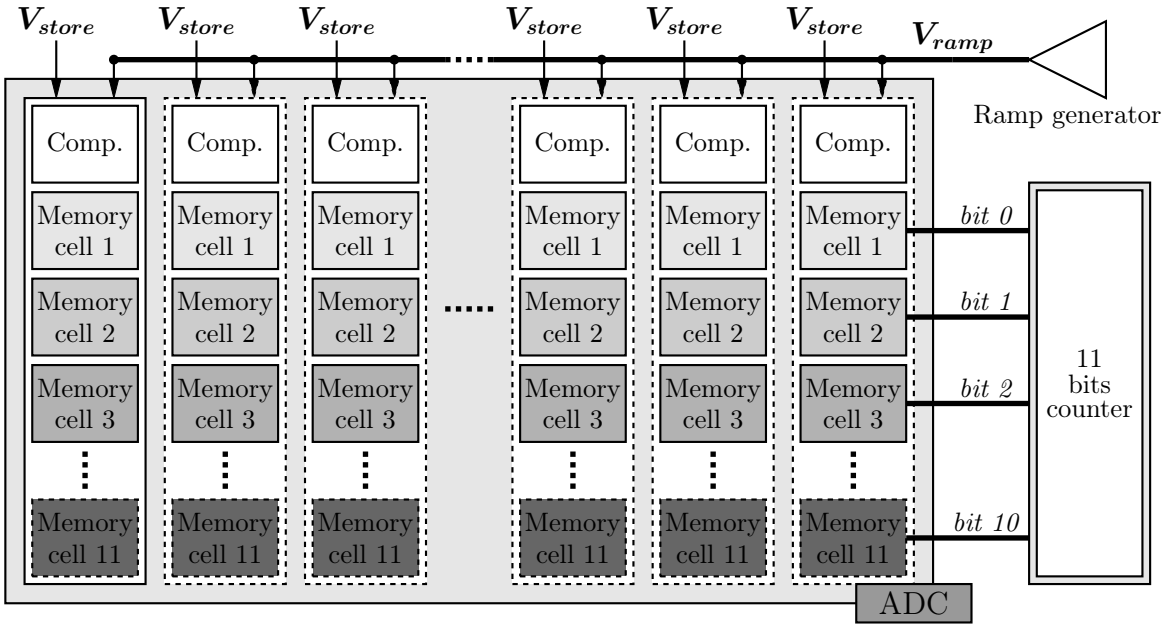


Figure 8.81 – Structure of a single channel Wilkinson ADC

8.9.3 Conversion principle

The principle of Wilkinson analog-to-digital conversion is illustrated in Figure 8.83. During the conversion, the stored voltage of each sampling capacitor V_{store} is continuously compared to a ramp voltage V_{ramp} using one comparator per sampling cell. The conversion begins by starting both the *ramp* and a digital counter. When the *ramp* reaches the *stored* voltage, the comparator fires, making the ADC memory cells

record the state of the counter. The basic architecture of this converter is shown in Figure 8.82.

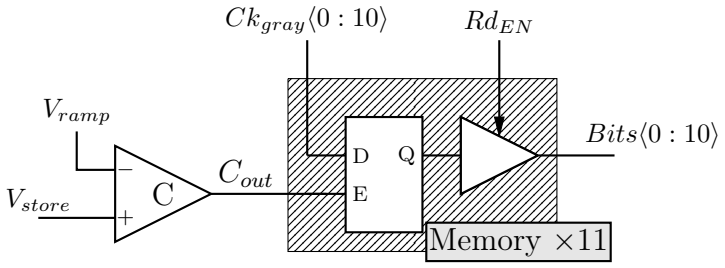


Figure 8.82 – Building blocks of an 11 bits Wilkinson converter. The comparator output latches the 11 bits of the counter in the memory cells.

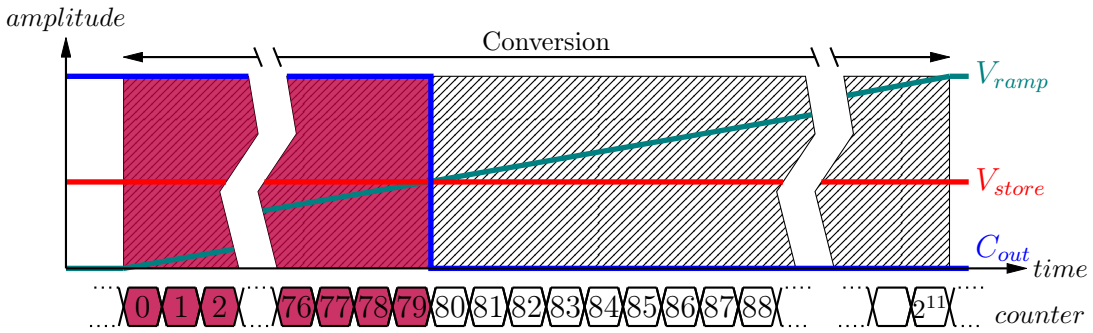


Figure 8.83 – Principle of the Wilkinson ADC conversion. During the conversion the stored voltage of each sampling capacitor V_{store} is continuously compared to a ramp voltage V_{ramp} using one comparator per sampling cell. The conversion begins by starting both the *ramp* and a digital counter. When the *ramp* reaches the *stored* voltage, the comparator fires, making the ADC memory cells record the state of the counter.

At t_0, the beginning time of conversion, a digital counter $c(t)$ is started together with a voltage ramp $V_{ramp}(t)$. If t_{conv} is the time required for the *ramp* to go from 0 to V_{DD}

$$V_{ramp}(t) = V_{DD} \times \frac{t - t_0}{t_{conv} - t_0} \; [t_0 \geq t \geq t_{conv}] \tag{8.41}$$

8.9.3.1 Conversion time

The time to complete a full conversion is given by the time required for the 11-bit ADC clock to count up to 2^{11}. In SamPic, the clock frequency f_{ck} is 1 GHz, so we have:

$$t_{conv} = \frac{2^{11}}{f_{ck}} \approx 2\mu s \tag{8.42}$$

8.9.3.2 Digitized value

The analog value V_{store} held on the capacitor of the sampling cell is continuously compared to the voltage ramp $V_{ramp}(t)$. When the voltage ramp V_{ramp} reaches the value stored on the sampling capacitor V_{store} the comparator output rises and latches the digital counter value. The value c of the counter when $V_{store} = V_{ramp}$ is given by:

$$c = \lceil \frac{V_{store}}{V_{DD}} 2^{11} \rceil \tag{8.43}$$

8.9.4 ADC comparator

The schematic of the ADC comparator is presented in Figure 8.84. It is a simple two stages CMOS comparator, the simplicity of the design was required in order to save space in the SamPic channels: there are 64 comparators per channel.

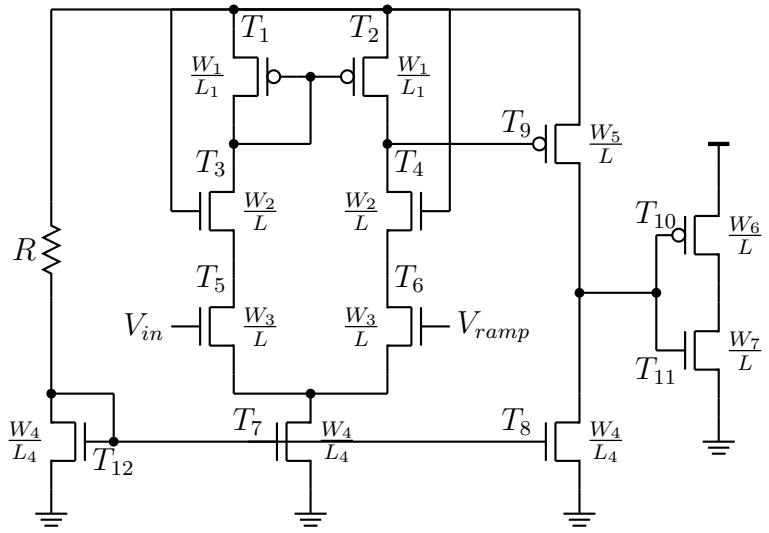


Figure 8.84 – Schematic of the ADC comparator with its external polarization resistor R. Note the presence of transistors T_3 and T_4 used to suppress the Miller effect affecting the input transistors $T5$ and T_6.

8.9.4.1 Comparator bias

The bias of the comparators is set via an external resistor R common to the full ASIC. The value can be modified for performance adjustments, *i.e.* gain/speed tradeoff, two modes in particular are used, a *high current* mode, where $R = 1.6M\Omega$, and a *low current* mode, where $R = 16M\Omega$.

8.9.4.2 Comparator gain

The gain must be high enough to distinguish a signal with an amplitude equal to the ADC Least Significant Bit ≈ 0.5 mV. A 0.5 mV signal difference should be amplified

into a signal large enough to be considered as a digital signal (1 V). This means a gain on the order of 2000.

The gain of the comparator simulated for different *store* voltages V_{store} is plotted in Figure 8.85. The gain is steady from 0.2 V to 1.4 V.

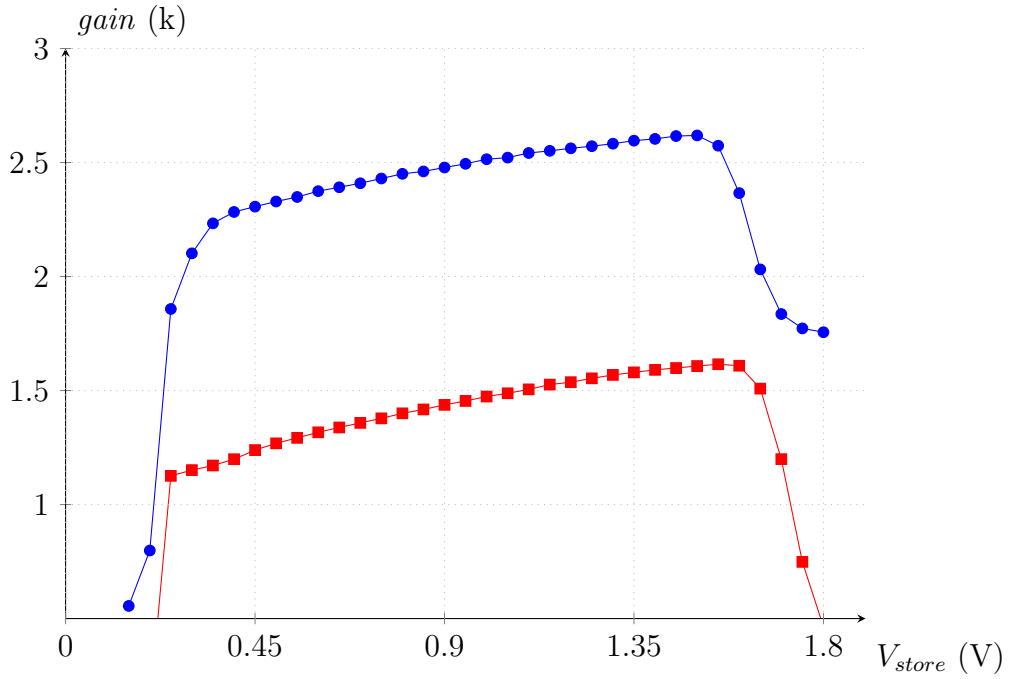


Figure 8.85 – Simualtion of the gain of the comparator as a function of the *stored* voltage done for *high current* ($R = 1.6M\Omega$) and *low current* ($R = 16M\Omega$)

8.9.4.3 Power consumption

The power consumption of one SamPic ADC comparator is given in Table 8.7.

Table 8.7 – Power consumption of one SamPic ADC comparator.

Currentmode	Consumption
high current ($R = 1.6M\Omega$)	$\sim 1\ \mu W$
low current ($R = 16M\Omega$)	$\sim 0.08\ \mu W$

8.9.4.4 Parasitic coupling

If we consider the input of the comparator directly connected to the sampling capacitor, we can simulate the effect of the ramp on the high impedance node of the sampling capacitor. During the *ramp* the amount of current i flowing through the transistor T_5 varies, so does the value of the parasitic capacitor of T_5, as shown in the simulation in Figure 8.86.

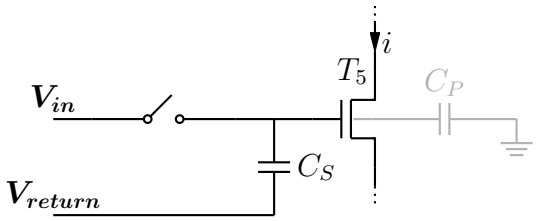


Figure 8.86 – Schematic of the equivalent circuit if the sampling cell was connected directly to the ADC comparator.

For noise and matching considerations, the size of the input transistors T_5 and T_6 (Figure 8.84), must be relatively large. Therefore, the variation of the parasitic capacitor of T_5 during the *ramp* is not negligible. Figure 8.87 shows a simulation of the variation of the imput parasitic capacitor on the sampling node in the case of direct connection to the comparator —●— or with an additional buffer —■—. We observe that the value of the parasitic capacitor is large (compared to the $35fF$ value of the sampling capacitor C_S). Additionally, this value varies strongly with V_{store}, which will make the bandwidth in the sampling cell hard to control.

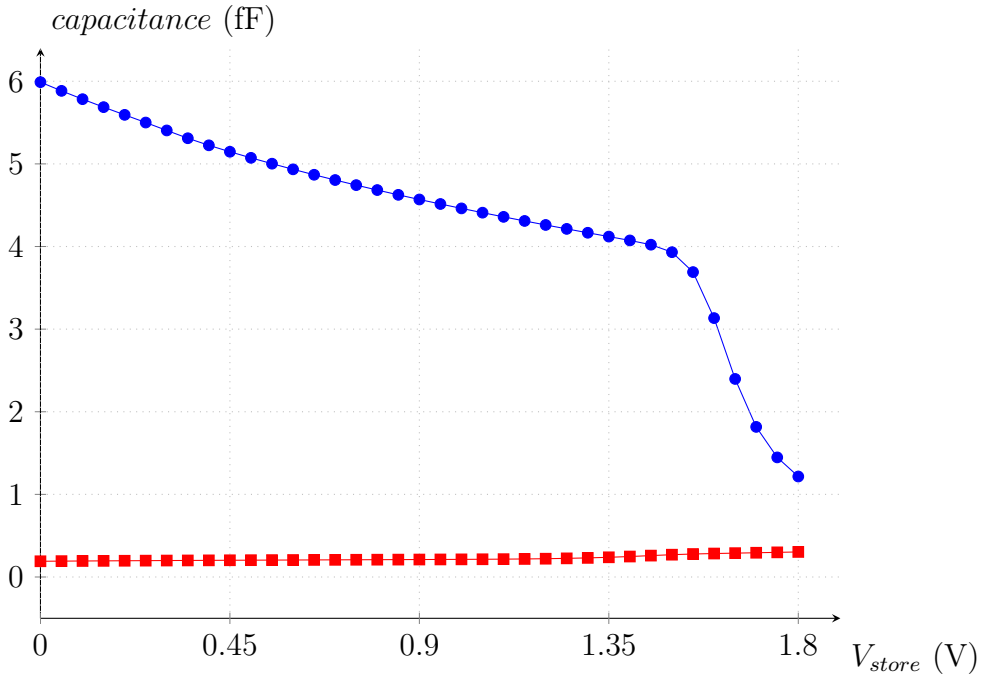


Figure 8.87 – Simulation of the variation of the input parasitic capacitor on the sampling node in the case of direct connection to the comparator —●— or with an additional buffer —■—. We can observe that the value of the parasitic capacitor is large (compared to the $35fF$ value of the sampling capacitor C_S). Additionally, this value varies strongly with V_{store} making the bandwidth hard to control. The use of the buffer reduces these effects by a factor more than 20.

An additional non-desirable effect of the parasitic capacitor of the input of the comparator is to increase the capacitor of the sampling cell reducing the bandwidth.

8.9.5 Input buffers

In order to remove the parasitic effects of the comparator input capacitor and the sampling cell *store* voltage, we add a simple source follower buffer in front, see Figure 8.88.

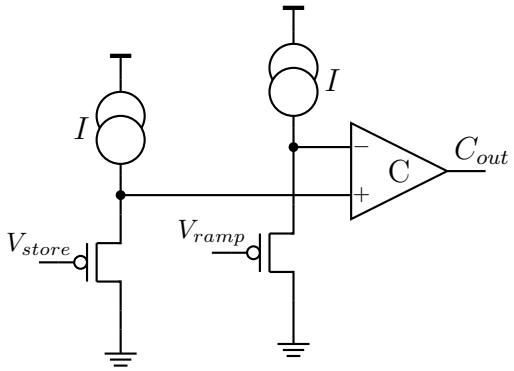


Figure 8.88 – Schematic of the comparator with input buffers.

The buffer is made as small as possible to reduce its parasitic capacitor on the sampling node (see Figure 8.87) and is biased by a fixed-intensity current. To conserve the symmetry during comparison operations, the same buffer is added for the *ramp* voltage.

8.9.6 Response time

The effect of response time of the comparator is illustrated in Figure 8.89. When V_{ramp} crosses V_{store} the comparator output fires after a propagation *delay*. If the *delay* of the comparator is not function of the *stored* voltage, or if the variation is smaller than a clock cycle, it can be considered as an offset during conversion. Otherwise it will add non-linearity to the ADC conversion.

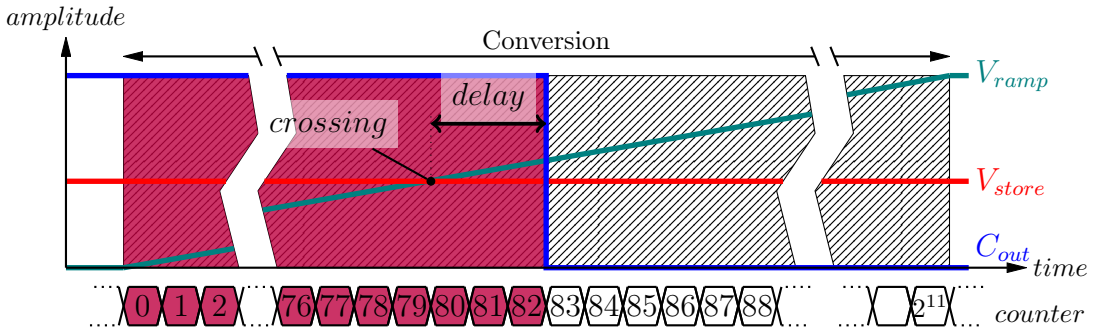


Figure 8.89 – Illustration of the effect of the delay of the comparator.

8.9.6.1 Simulation results

A simulation of the response time of the comparator with its input buffers is reported in Figure 8.90.

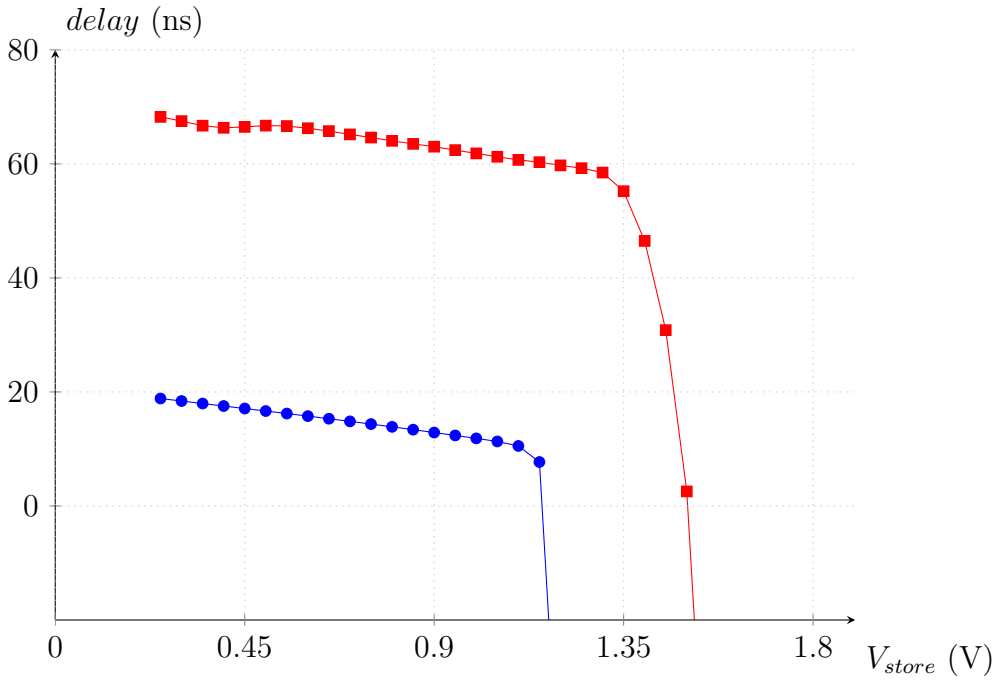


Figure 8.90 – Simulation of the response time of the comparator function of the *store* voltage, done for *high current* ($R = 1.6M\Omega$) and*low current* ($R = 16M\Omega$).

Unfortunately, the *delay* of the comparator is not steady as a function of V_{store}, it is however a linear function of the sampled voltage with a slope of about $-10^{ns}/_{V}$ between 0.25 V and 1.1 V, so that it can be seen as a change of the overall channel gain.

8.9.6.2 Monte Carlo simulation

A simulation of the comparator response time *delay* variations for process and mismatch variations is presented in Figure 8.91. This simulation reflects the result of non-uniform transistor manufacturing in an ASIC and from ASIC-to-ASIC, on their electrical and physical parameters. The simulation is made with 600 entries. The standard deviation is simulated to be about 3 ns. This corresponds to a contribution to conversion offsets from cell to cell of 3-bits with a conversion clock frequency $f_{wilk} = 1$ GHz.

These deterministic offsets will produced fix offsets in the converted ADC value from cell-to-cell and from ASIC-to-ASIC. However, they can easily be calibrated out permanently and automatically substracted after readout in the data acquisition system.

8.9.6.3 Temperature dependence

The temperature dependence of the *delay* of the comparator is presented in Figure 8.92 for temperature going from $T = 0°C$ to $T = 75°C$.

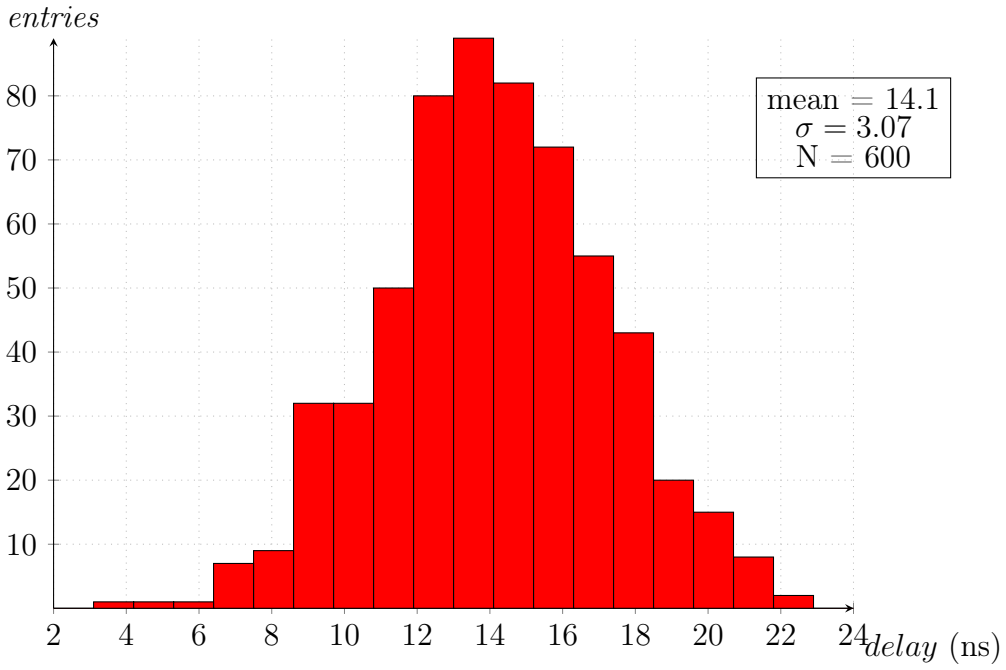


Figure 8.91 – Monte Carlo simulation of the comparator delay done considering both process and mismatch variations for 600 events.

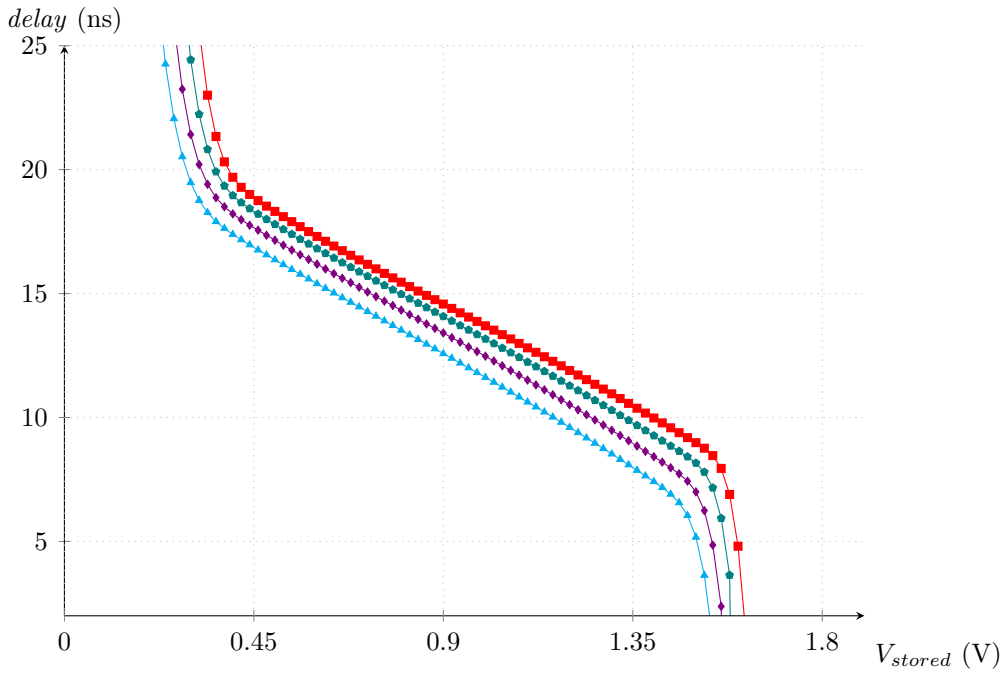


Figure 8.92 – Simulation of the delay of the comparator versus the stored voltage V_{store} on the switch, for the temperature $T = 0°C$ —●—, $T = 25°C$ —◆—, $T = 50°C$ —■—, $T = 75°C$ —▲—.

The dependence observed if of 2 ns / $75°C$, which means that the conversion is

stable within 1 LSB for temperature variations of $37°C$.

8.9.7 Comparator design

The layout of the comparator used for the analog to digital conversion is shown in Figure 8.93. As it was illustrated in the simulation shown in Figure 8.91, the delay response of the comparator varies with physical and electrical mismatches of its PMOS and NMOS transistors. In order to reduce these effects as much as possible a carefull attention was given to the layout of the comparator input, the *differential input pair* in particular, in order to maximize symmetry in the design and avoid as much as possible any offsets from comparator to comparator across the 64 cells of the 16 channels of SamPic.

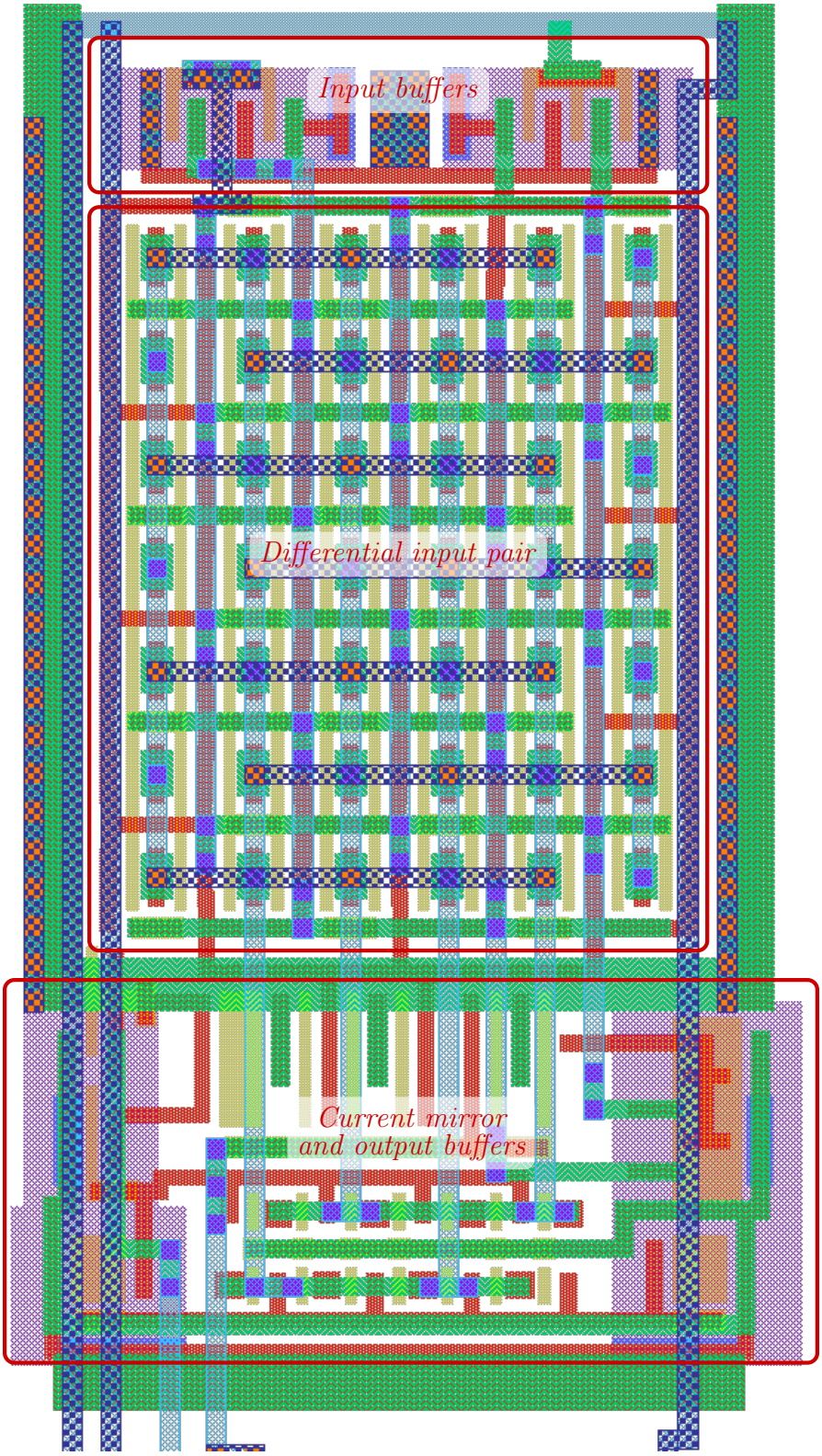


Figure 8.93 – Layout of the comparator.

8.9.8 Power and noise reduction

In order to reduce the power and noise created by the $11 \times 64 \times 16 = 11264$ memory cells during the conversion (11 latches in 64 cells of 16 channels), we avoid making the latch transparent to the Gray counter before the comparator triggering. Instead of using the comparator rising edge for latching the counter value, it is used to form a pulse to track the counter value for a small amount of time. As a result the latched value of the counter is offset by the pulse duration *delay* compared to the real time of ramp to stored value crossing.

Using this technique, the reduction in power consumption, PRF (power reduction factor), for a 11-bit conversion at 1 GHz with a *delay* of 1 ns can be as high as:

$$PRF = \frac{conversiontime}{delay} = \frac{2^{11}}{1Ghz * 1ns} = 2048 \tag{8.44}$$

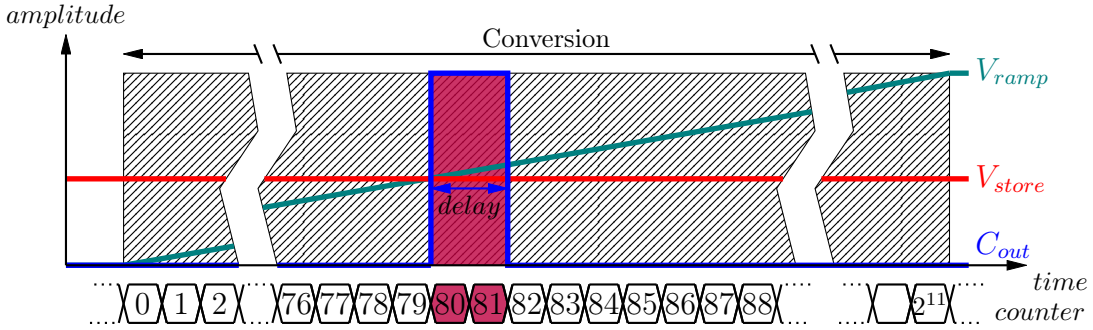


Figure 8.94 – Principle of the ADC conversion using the power reduction technique.

The use of a small pulse to latch *Enable* the Wilkinson clock has the advantage of reducing the commuting noise in the memory cell: we can have up to 11264 latches commuting during a conversion. The pulse reduces the commuting time, hence the noise. The pulse duration is approximately $1ns$. See Figure 8.94.

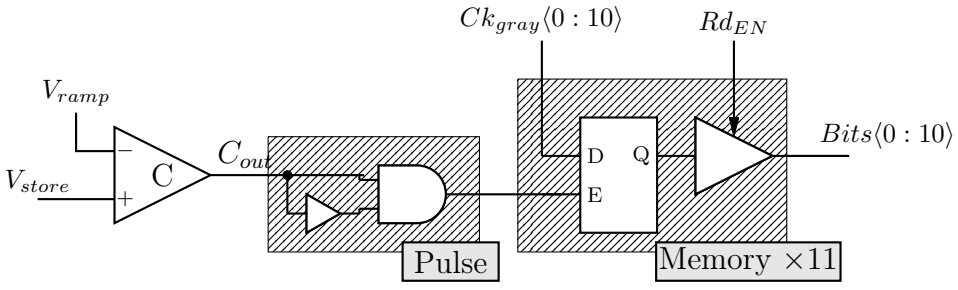


Figure 8.95 – Building blocks of one ADC cell

8.9.9 Ramp generator

The ramp generator is shown in Figure 8.96. It consist of a capacitor C charged by a low-voltage cascode current mirror. When the switch S is *open*, the externally applied

current I_{pol} flows to the capacitor and charges it up. If I_{pol} is held at a constant value, the voltage across the capacitor grows linearly and is used as the *ramp* voltage for the ADC. Practically, I_{pol} is a temperature stabilized tunable current source.

The time domain simulation of the voltage ramp is shown in Figure 8.97. The linearity of the *ramp* directly sets the linearity of the ADC conversion.

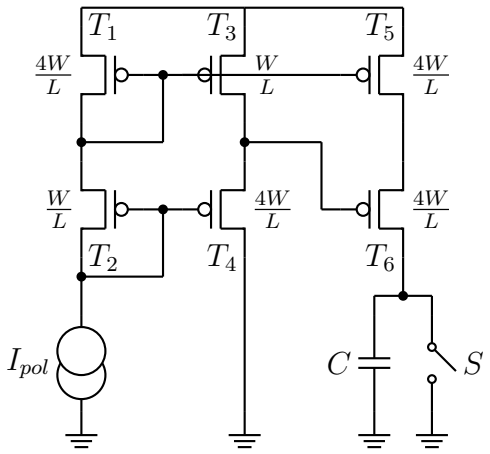


Figure 8.96 – Schematic of the ramp generator structure. The current source I_{pol} is external and compensated for temperature variations.

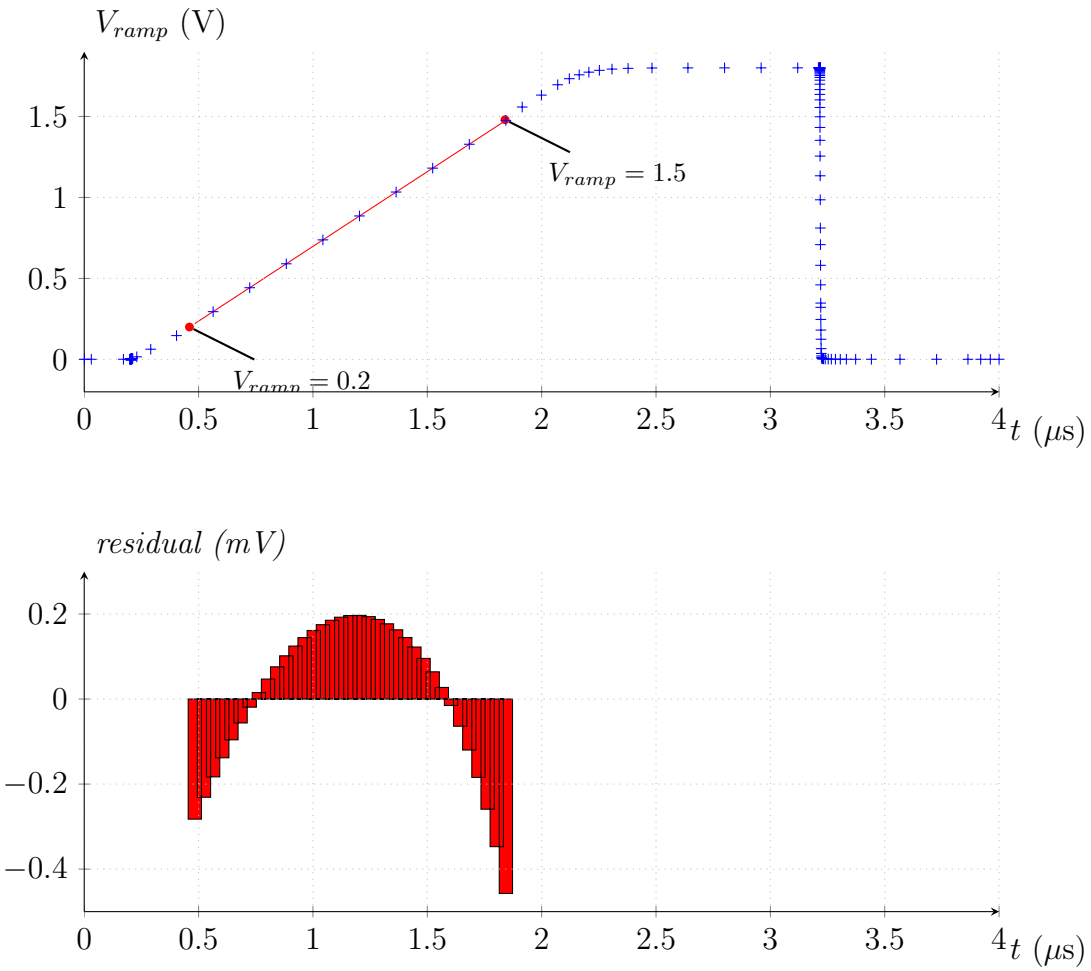


Figure 8.97 – Time-domain simulation of the voltage ramp V_{ramp}. The ramp runs from 0 to 1.8 Volt. It is fitted in its most linear region (between 0.2 and 1.5V) and the residues show a integral non-linearity of 0.5 mV. It corresponds to 0.8bit of non-linearity coding the same range on 11-bit.

8.9.10 Gray counter

The analog ramp causes comparator outputs to change state asynchronously with respect to the counter clock. If a standard binary counter was used, the asychronous comparator output has a non-zero probability to latch the changing output of the counter, which can have several bits changing simultaneously between two consecutive codes, resulting in unpredictible results. For example, if the comparator triggers during the 7-to-8 transition of the counter, a wrong 15 value can be latched in the memory (see Table 8.8).

Instead of this, a Gray code counter is used, which has the property that only one bit changes going from one state to the next. This reduce the error to the least significant bit [9],[10].

Table 8.8 – 4-bit binary and Gray code representation.

Value	Binary	Gray
0	0000	0000
1	0001	0001
2	0010	0011
3	0011	0010
4	0100	0110
5	0101	0111
6	0110	0101
7	0111	0100
8	1000	1100
9	1001	1101
10	1010	1111
11	1011	1110
12	1100	1010
13	1101	1011
14	1110	1001
15	1111	1000

8.9.11 Digital memory cells

The digital memory cells are shown in Figure 8.95. They are made of one latch an one tri-state buffer, permitting the readout of the cell.

8.10 Triggering

Each Sampic channel has capabilities for internal and external triggering. The trigger stops the sampling when a event of interest happens and stores it on the SCA for future conversion and readout. For each sampling cell of the SamPic channels this sequence is controlled by a *Trigger cell* . See Figure 8.98.

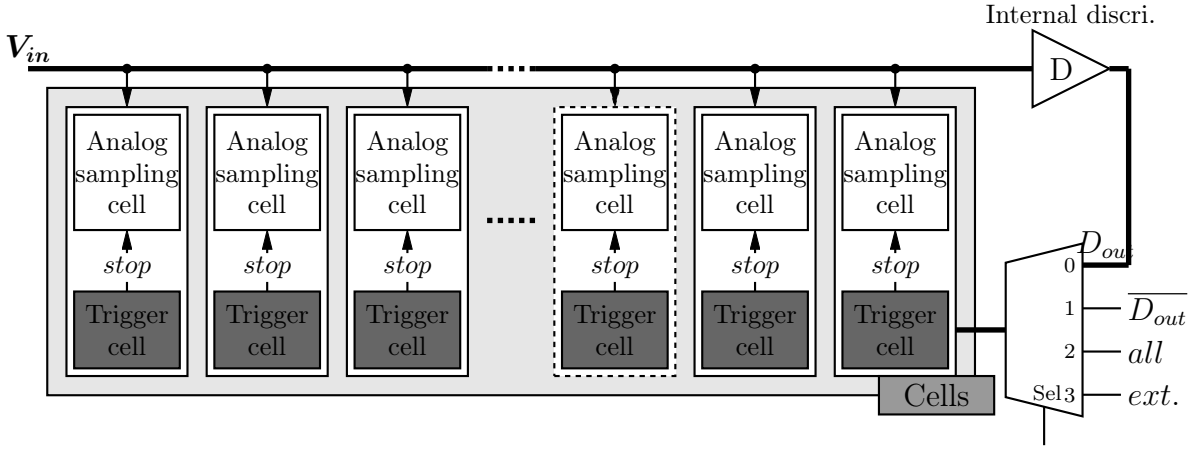


Figure 8.98 – Triggering path and options inside a SamPic channel.

Four triggering options are available for each of SamPic's channels:

Table 8.9 – Trigger mode options.

Option	Trigger mode
0	external
1	internal on rising edge
2	internal on falling edge
3	central trigger

The trigger option for each channel is set internally using the channel registers, it is programmed externally using a SPI-like interface.

8.10.1 The trigger cell

There is one trigger cell per sampling cell in the sampling capacitor array (see Figure 8.98). This cell plays three roles in the sampling process:

Freeze The principal goal of the trigger cell is to stop the DLL sampling windows for each sampling cell in case of a trigger.

Buffering An additional goal is to locally rebuffer and control the width of the sampling window in order to avoid sampling time distortions.

Region Of Interest The trigger cells also register the time a which the trigger signal happened with respect to the sampling window: it flashes the position of the sampling windows at the time of the trigger (see Figure 8.108).

8.10.2 Trigger modes

8.10.2.1 External trigger

The external trigger mode allows to externally stop the sampling in all the SamPic channels set in *mode 0* (see Table 8.9). The *external trigger* signal is a SamPic input signal. The operation mode is shown in Figure 8.99.

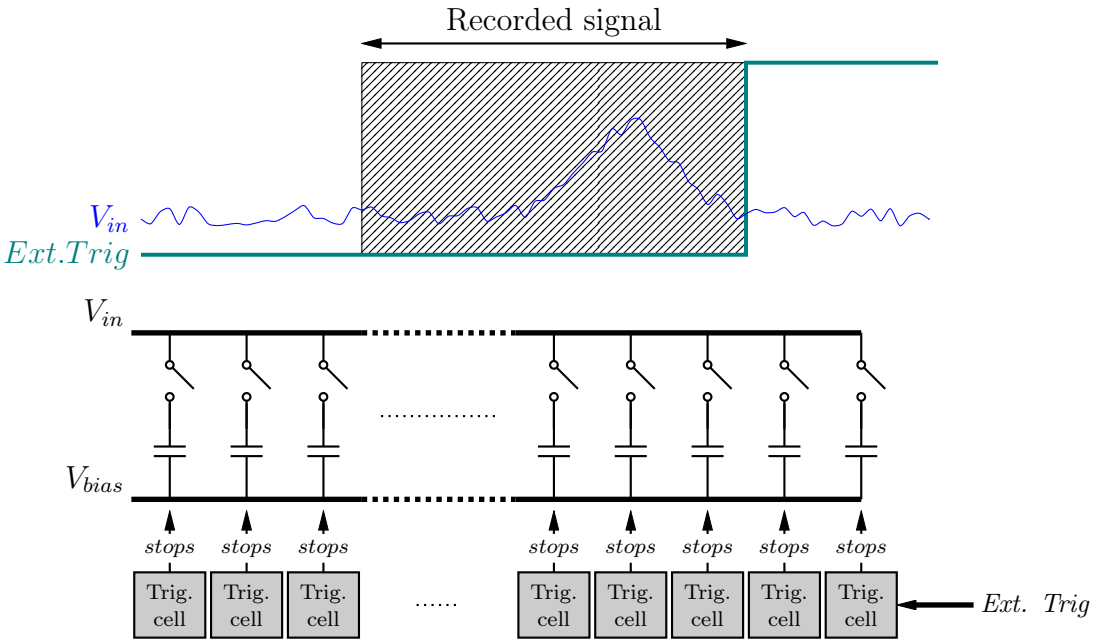


Figure 8.99 – In the external triggering mode, the trigger signal is sent externally to each SamPic's channel. The 64 data points that where taken before the trigger signal occurred are stored.

8.10.2.2 Internal discriminator on rising edge

For the SamPic channels set in *mode 1* (see Table 8.9) the output of the channel internal discriminator stops the sampling when the input signal V_{in} rises above the threshold voltage V_{thresh}. The threshold voltage can be set using the channel internal DAC or an externally applied threshold level. This operation mode is shown in Figure 8.100.

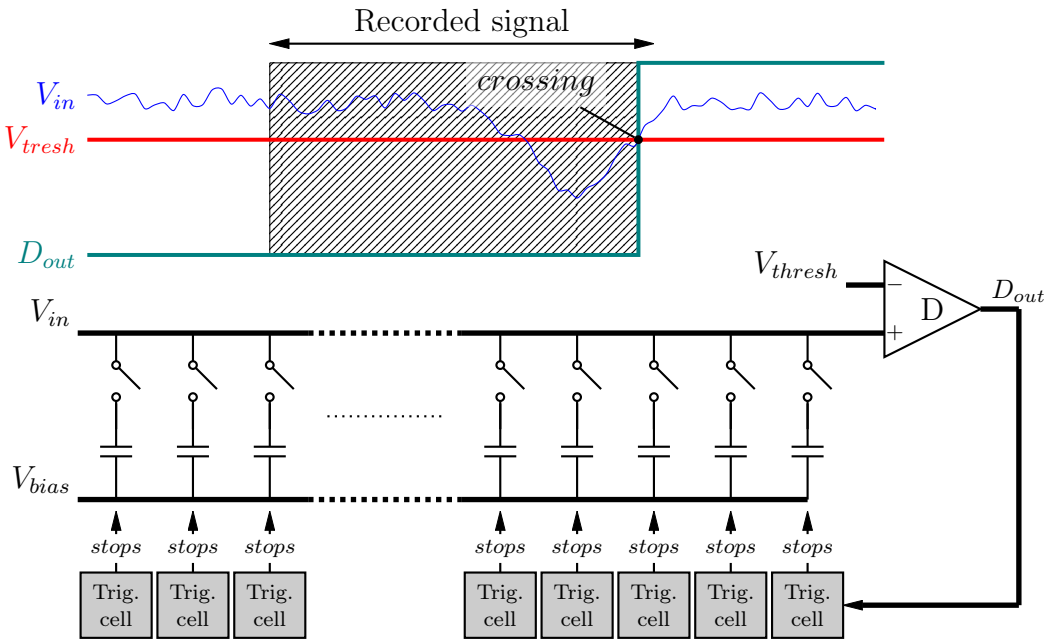


Figure 8.100 – In the internal trigger mode, the input signal is continuously compared to programmable threshold using the channel discriminator. In the rising edge mode when V_{in} rises above V_{thresh}, the discriminator fires and stops the sampling.

8.10.2.3 Internal discriminator on falling edge

For the SamPic channels set in *mode 2* (see Table 8.9) the output of the channel internal discriminator stops the sampling when the input signal V_{in} falls below the threshold voltage V_{thresh}. The threshold voltage can be set using the channel internal DAC or an externally applied threshold level. This operation mode is described in Figure 8.101.

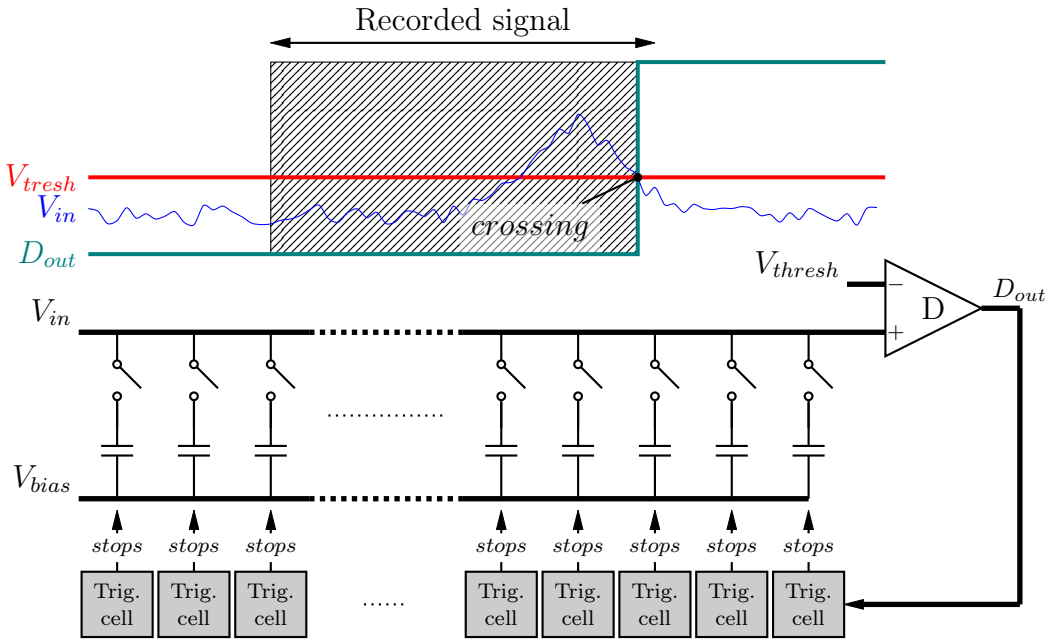


Figure 8.101 – In the internal trigger mode, the input signal is continuously compared to the programmable threshold using the channel discriminator. In the falling edge mode, when V_{in} falls below V_{thresh}, the discriminator fires and stops the sampling.

8.10.2.4 SamPic central trigger

For the SamPic channels set in *mode 3* (see Table 8.9) the sampling is stopped when the discriminator output of all the channels have fired. This trigger is made by a logical OR of the output of all the enabled internal discriminators of the SamPic chip. This operation mode is shown in Figure 8.102.

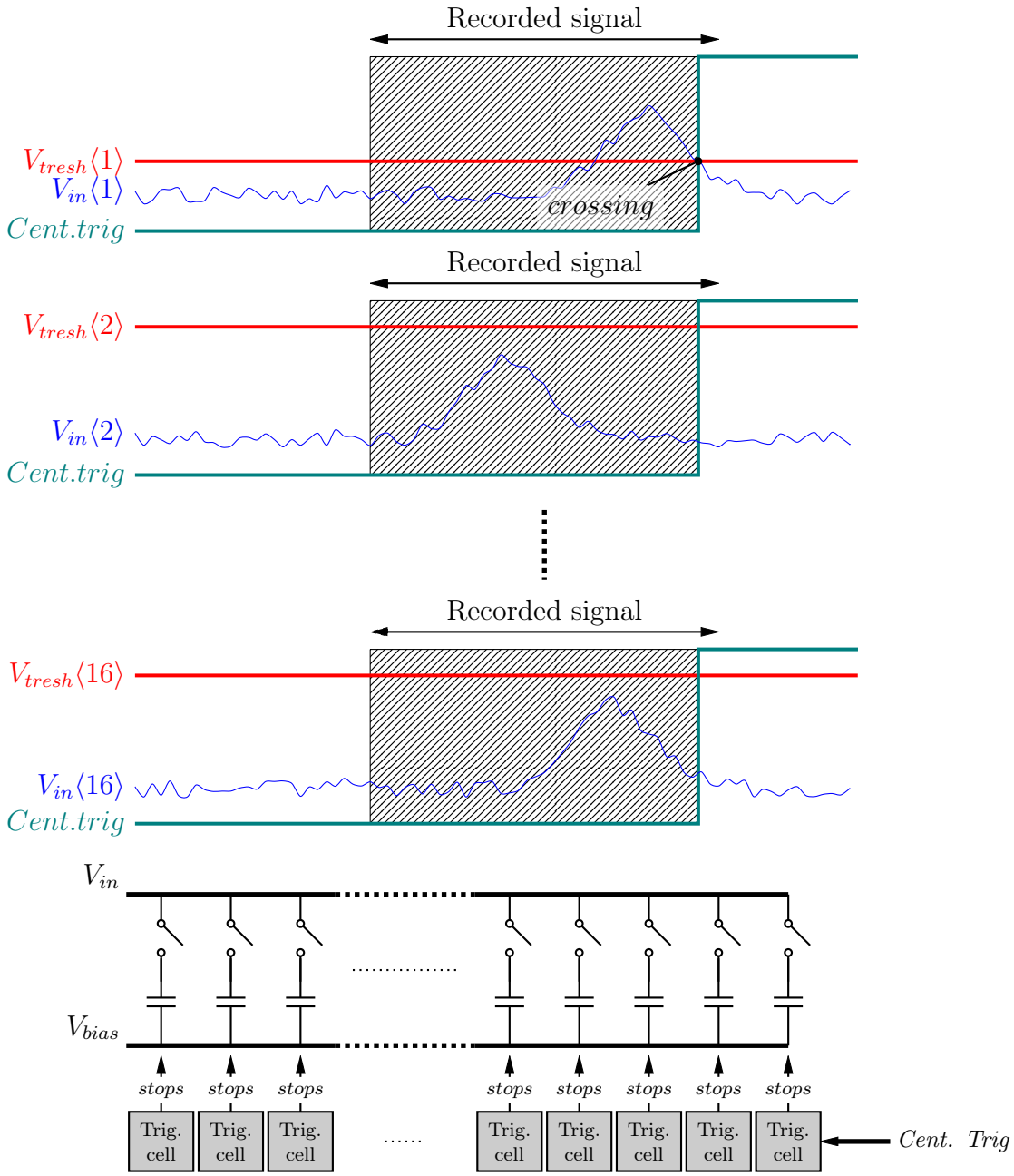


Figure 8.102 – In the central trigger mode if a discriminator fires it stops the sampling for all the channels. In this Figure the channel 1 discriminator fires and stops the sampling in all channels.

8.10.3 Channel triggering

8.10.3.1 Post trigger

The post trigger principle is illustrated in Figure 8.103. It allows for additional sampling time before generating the real trigger signal. For example to record the tail of a positive input pulse triggered on the rising edge, record double pulses or other features. Inside SamPic the post-trig *delay* can be chosen between 3 different values:

Table 8.10 – Trigger delay options

Option	Delay
0	no delay
1	600ps
2	1300ps

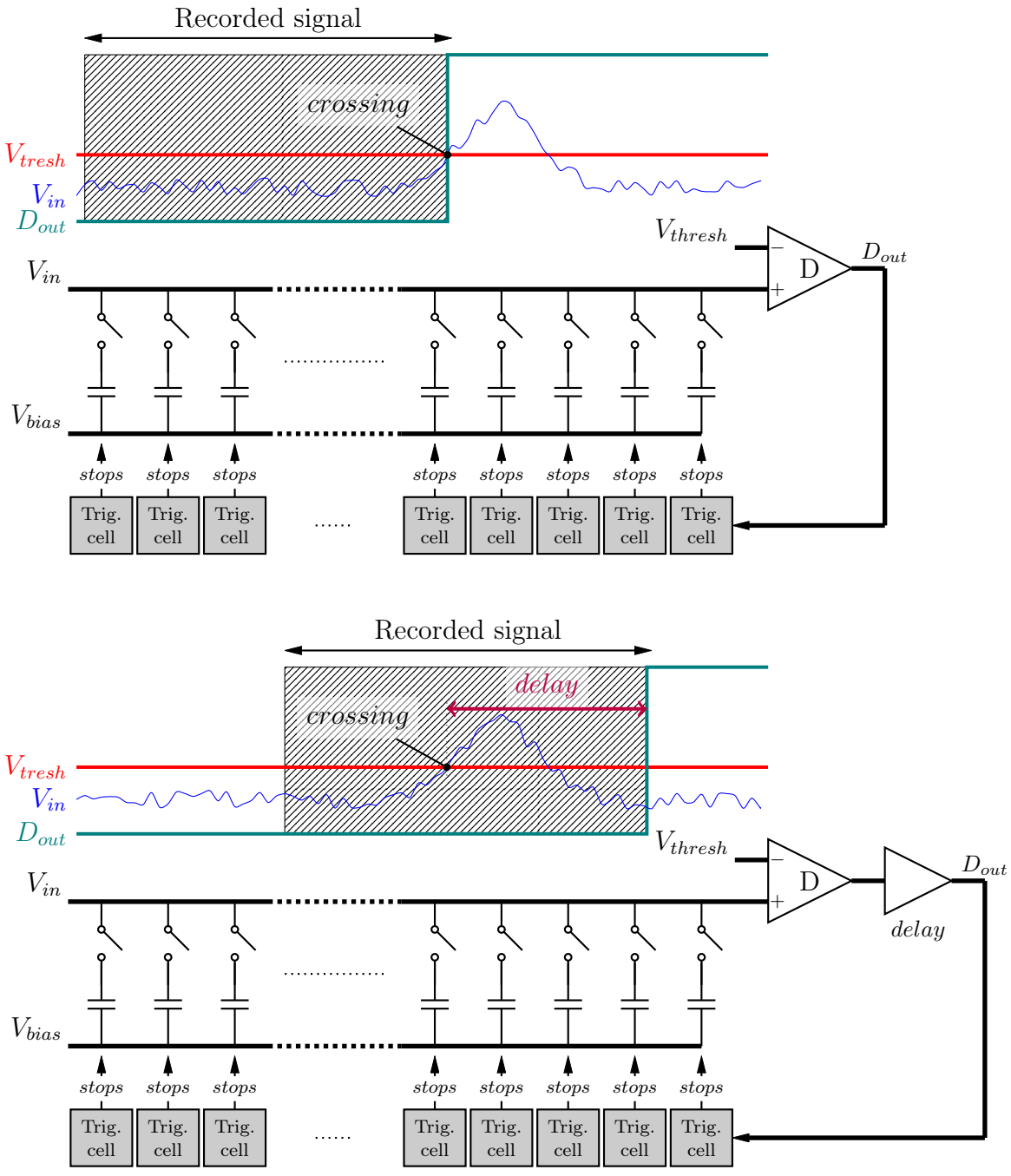
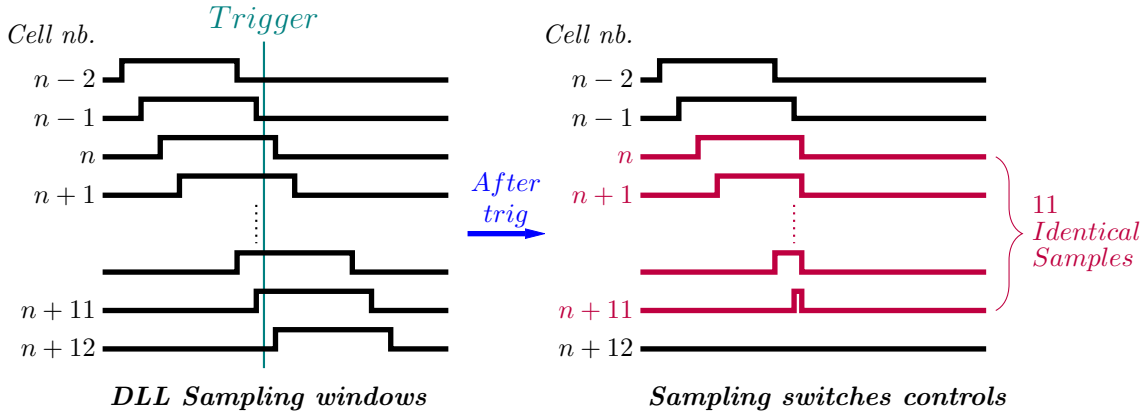


Figure 8.103 – Description of the role of the post-trigger. A controllable delay can be added after the discriminator output in order to record additional signal after a crossing. This can be interesting in the case of a positive pulse and the trigger set on the rising edge (as shown on the picture). If no delay was added, the falling edge part of the pulse would not have been recorded.

8.10.3.2 Dead cell removal

By construction of the DLL, at each sampling step, a given number n of cells are connected to the input of the channel. If stopped abruptly by a trigger event, the same input signal would be sampled on these n sampling cells. A smart trigger stopping scheme inside each channel removes this redundancy by taking $n-1$ samples after the trigger and therefore catching 64 distinct samples.

Case 1: Trigger signal stops the sampling windows unconditionally.



Case 2: Trigger ends the sampling, after the end of the write sequence.

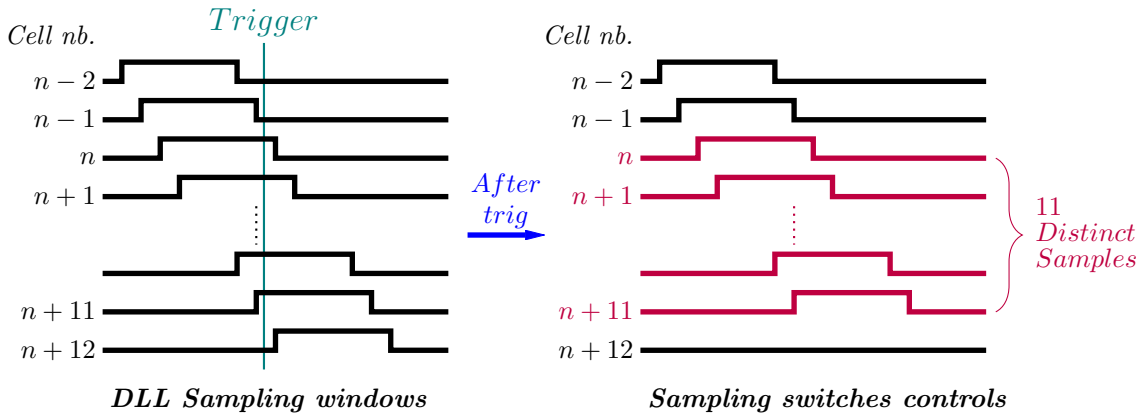


Figure 8.104 – Description of the principle of dead cells removal. In the first case the sampling array is stopped unconditionally when a trigger signal is sent to the sampling array. Due to the width of the sampling signals equal to 12 times the sampling step, it produces 11 sampling windows ending at the same time, therefore we are recording 11 times the same input signal in the sampling capacitor array. The stopping scheme used in SamPic is shown in the case number 2. When the trigger signal fires the sampling signal of the cells in write mode are allowed to finish while all the other cells are held in hold mode. All samples in the SCA after triggering are distinct.

This operation is done inside the trigger cell. The structure used for that purpose is the following:

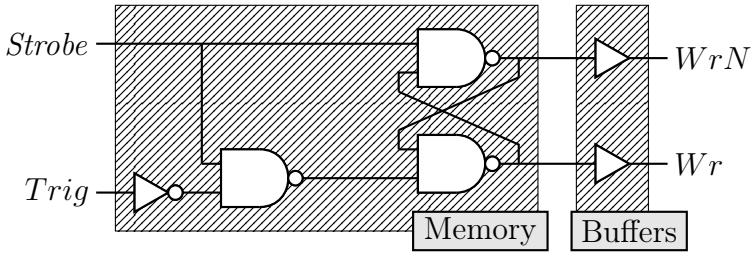


Figure 8.105 – Schematic of the circuit used for dead cell removal inside the trigger cells.

The principle is the following: in normal operating mode the $Trig$ signal is low. The equivalent circuit is:

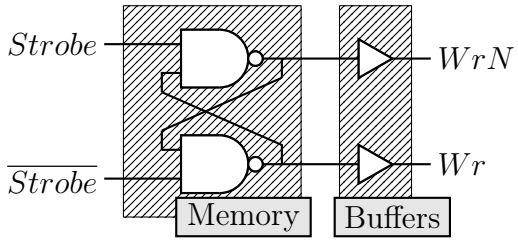


Figure 8.106 – Equivalent circuit in sampling mode ($Trig = 0$).

So in normal operating mode:

$$\begin{aligned} Wr &= Strobe \\ WrN &= \overline{Strobe} \end{aligned} \tag{8.45}$$

When $Trig$ goes high ($Trig = 1$), the equivalent circuit is:

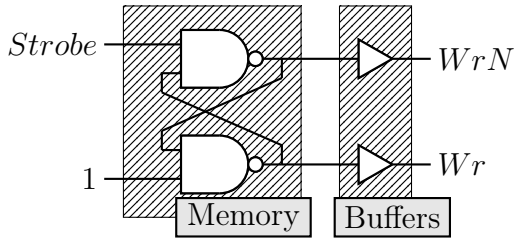


Figure 8.107 – Equivalent circuit in trigger mode ($Trig = 1$).

So in *trigger* mode:

- If $Trig = 1$ and $Strobe = 0$, Wr stays at 0
- If $Trig = 1$ and $Strobe = 1$, $Wr = Strobe$. When $Strobe$ goes back to 0, Wr stays at 0.

8.10.4 Region of interest

The principle of the region-of-interest, as it was described in Figure 8.9, is to flag the cells in SamPic's channels where there is data of interest. Typically we expect this data to be close to the triggering time. Indeed, in most cases, the trigger is set when the input signal V_{in} crosses the threshold voltage V_{tresh}.

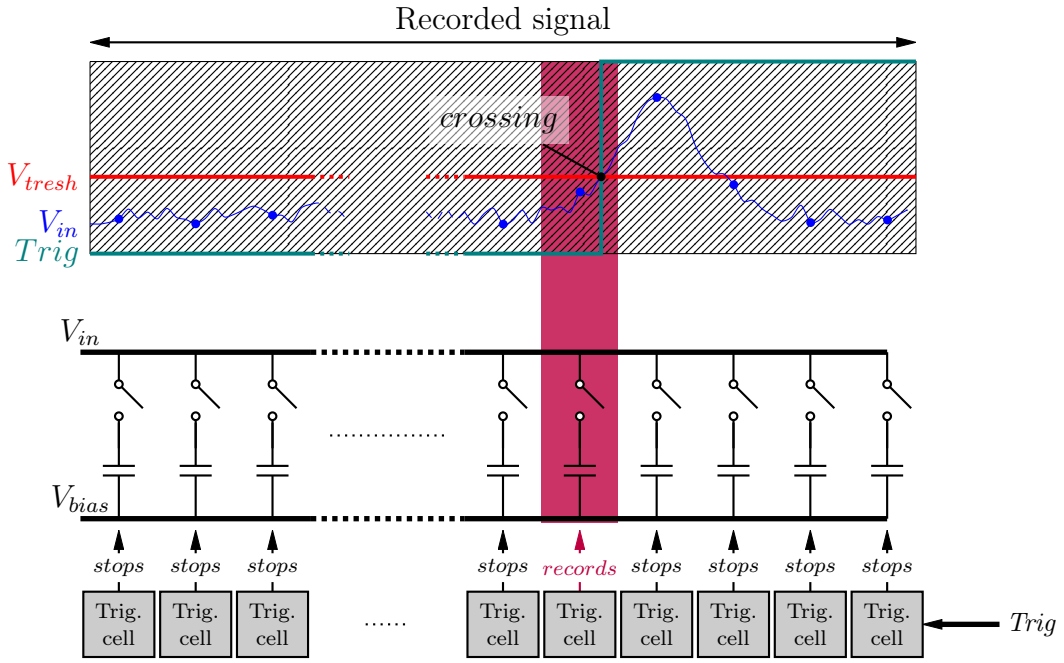


Figure 8.108 – Principle of the determination of the region of interest. When the trigger signal fires (here on the rising edge of the input signal), the *Trigger cell*, which was in sampling mode records the *Trigger* hit (in purple). This way, the position of the trigger inside the sampling array is known.

The determination principle of the region-of-interest is shown in Figure 8.108. First, the position of the *Trigger* hit with respect to the sampling array is recorded. This position is encoded and a programmable offset is added to determine the cell index at which the sequential readout will start. At last, it will be possible to read only a few cells corresponding to the interesting part of the signal [11].

8.10.4.1 Trigger hit recording and encoding

In order to record the position of the *Trigger* hit inside the sampling array, we use a similar technique to the one use for dead cells removal. We are recording the cells that were in *Write* mode when the *Trigger* signal occurred. The circuit used is a simple $\overline{RS}$ latch (see Figure 8.109). At any time there are always 12 cells in *Write* mode, therefore when the *Trigger* happens, 12 consecutive *Hits* are recorded in the trigger cells.

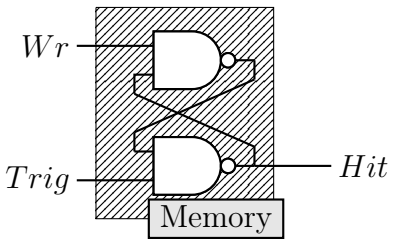


Figure 8.109 – Schematic of the circuit used for the *Trigger* hit recording inside the trigger cells.

The precise position of the *Trigger* hit with respect to the sampling cells is found by searching for the *011* pattern for 3 consecutive *Hit* values in the trigger cell. The *011* pattern detection allows removal of bubbles and fake hits.

This *Trigger* hit position, which can take 64 different values is then coded onto 6 bits and sent out during a channel readout. This process is shown in Figure 8.110.

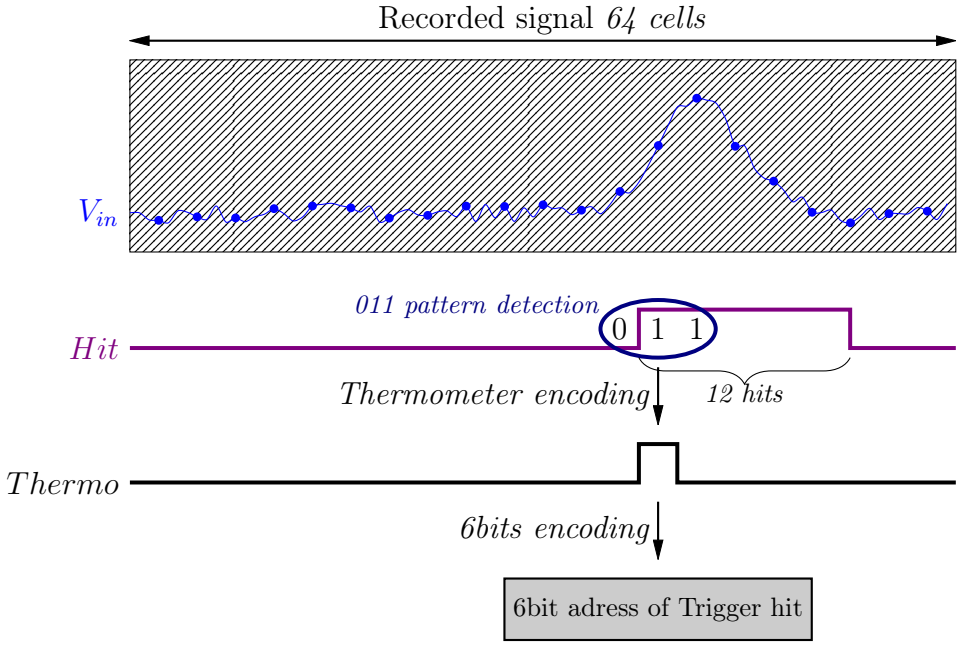


Figure 8.110 – Principle of the *Trigger* encoding as implemented in each SamPic's channels.

8.10.5 Internal discriminator

In SamPic, the triggering of each channel can be done by applying an external signal or using the internal discriminator present in each channel.

8.10.5.1 Specifications

Speed The discriminator must be fast: in SamPic, at 10 GSa/s, we only have 6.4 ns of sampling depth, if the discriminator latency is higher than this time we loose the event of interest because the sampling capacitor array will have been written with new data by the time the trigger stops the sampling.

Gain The discriminator must be high gain: we would like to be able to discriminate 1 ns FWHM pulses 1 mV above the threshold. That's a gain of:

$$G = \frac{1.8V}{1mV} = 1800 \tag{8.46}$$

At a bandwidth of:

$$BW = \frac{.35}{500ps} = 700MHz \tag{8.47}$$

Noise The noise of the discriminator should be below the LSB value of the channel ADC.

Power Power consumption of the discriminator is not critical.

8.10.5.2 Design

The schematic of the internal channel discriminator is described in Figure 8.111. It consists of four identical cascaded differential pairs, each achieveing a static gain of 3000, which adds up to 82dB when cascaded as shown on Figure 8.112. This type of structure is known to offer the best gain/power consumption trade-off.

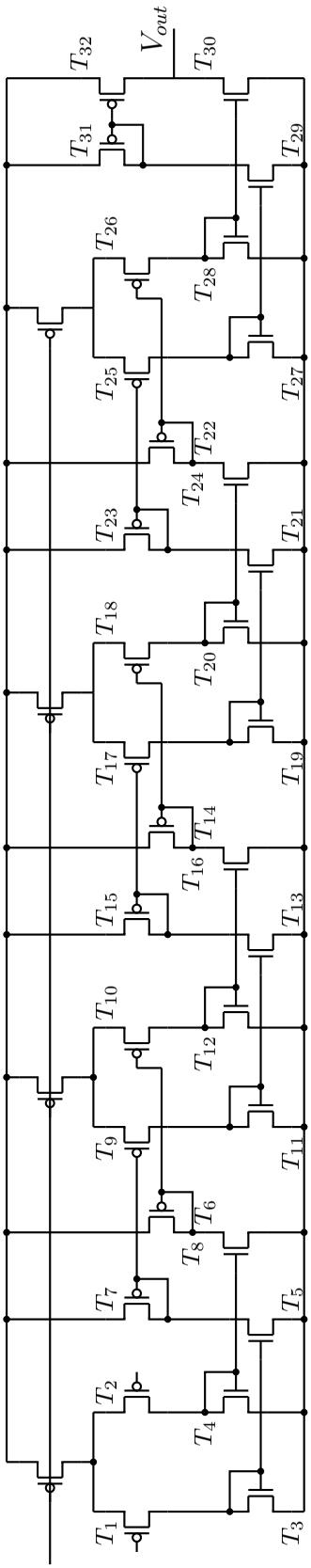


Figure 8.111 – Schematic of the four-stage discriminator.

Gain A plot of the gain of the discriminator is shown in Figure 8.112.

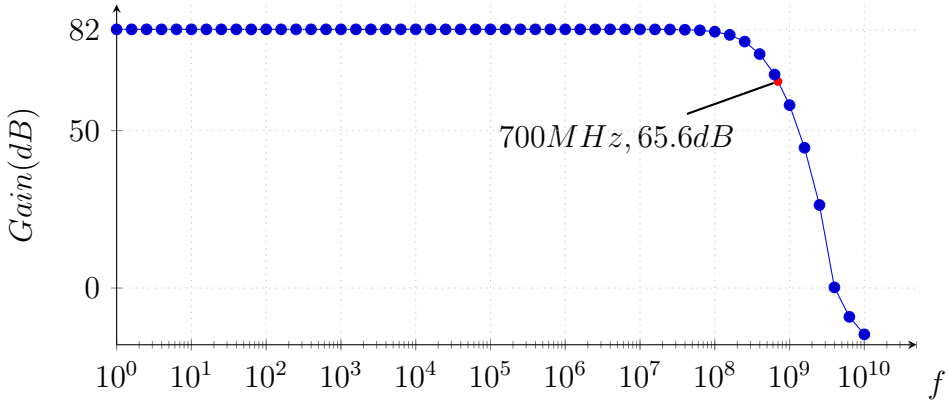


Figure 8.112 – AC Simulation of the gain of the discriminator versus the input frequency.

The gain of the discriminator at 700MHz is simulated to be 65.6dB = 1900. Therefore, the criteria to be able to discriminate 1ns FWHM input pulses 1mV above the threshold is met. This is also verified by transient SPECTRE simulation of the discriminator.

Speed In order to be able to have such a high gain at 700MHz we designed a four-stage discriminator (see Figure 8.111). Due to its important number of stages, the discriminator has a relatively high delay. The simulation of the delay function of the threshold voltage V_{thresh} is shown in Figure 8.113.

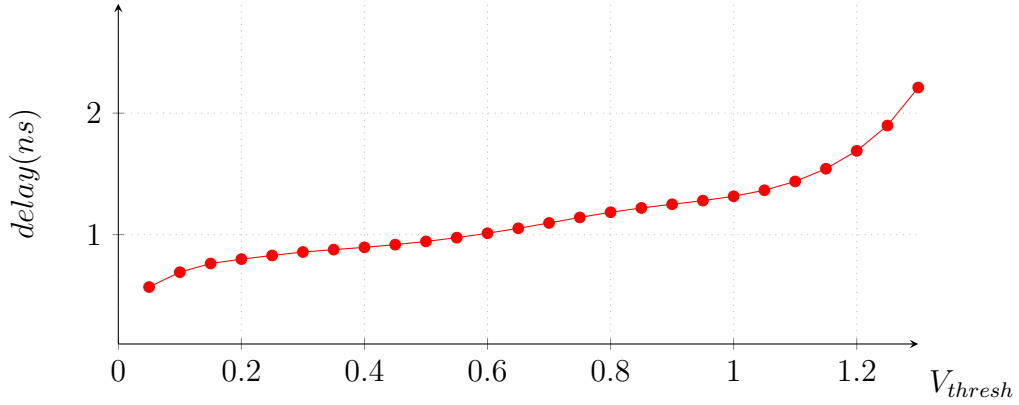


Figure 8.113 – Simulation of the delay of the discriminator versus the threshold voltage V_{thresh}.

This delay, as long as it is smaller than the sampling capacitor array buffering time, is not critical for the final timing extraction as it is performed using the sampling data only.

8.11 Time stamp

In order to keep track of the coarse time of the sampling window, a 12-bit Gray counter clocked with SamPic reference clock Ck_{ref} is implemented. When one of the SamPic channel triggers, the state of the counter is recorded in the channel. In order to resolve metastability errors, which would happen if the trigger fires at the exact time of the reference clock rising edge, both the timestamp and the timestamp delayed by $1/2 Ck_{ref}$ (Timestamp A and Timestamp B) are recorded in the channel. This principle is illustrated in Figure 8.114.

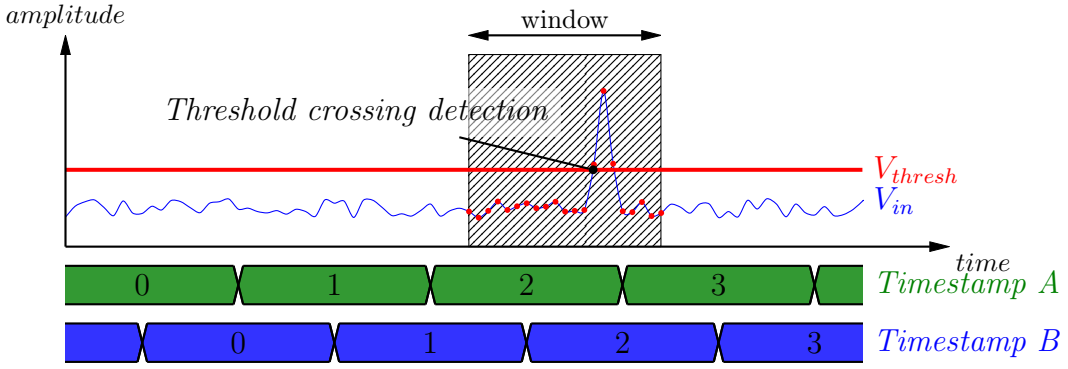


Figure 8.114 – The role of the timestamp counter is to keep track of the coarse time inside SamPic. At the time of the trigger the value of the counter is latched inside the channel and readout along with the data sample. A tag of two counters shifted by half the clock period allows for metastability removal.

8.12 Chip Conversion and Readout

The electronics managing the readout and conversion of SamPic has been designed to optimize deadtime while keeping versatility. This design, half located inside the channels, half in a common block, is using the digital standard cells from the AMS library and will not be described here. Instead, the rules it follows and some sample chronograms are given.

The readout and conversion of all the SamPic channels are controlled by external signals. The conversion and readout sequence follows this simple rule scheme:

Rule 1 All the SamPic channels that are not triggered are tacking data (sampling).
The signal "ChannelToConv" warns the user that one or more channel is triggered

Rule 2 All triggered channels that are not holding data to read are converted simultaneously when the user sends back the "Conv" signal.
The signal "ChannelToRead" warns the user that data are waiting for readout

Rule 3 All converted channels are readout sequentially when the user sends back the "Read" signal. The operation is sequenced by the read clock RCk.

The data is collected on 12 LVDS pairs. For each channel, the structure of the readout data is the following:

- First 12 bits: 6 bits holding the channel address and 6 bits holding the *region of interest* (see Figure 8.110).
- Then 12 bits holding TimeStamp A, the value of the 12 bit coarse counter at the time of the trigger.
- 12 bits holding TimeStamp B = TimeStamp A+$Ck_{ref}/2$, the value of the 12 bit coarse counter shifted by half a clock period allowing to solve potential metastability issues.
- 12×n bits of data, $n \in [1, 64]$ user defined, corresponding to n samples starting from a fixed offset in the sampling cell array or from the *region of interest* (see Figure 8.110).

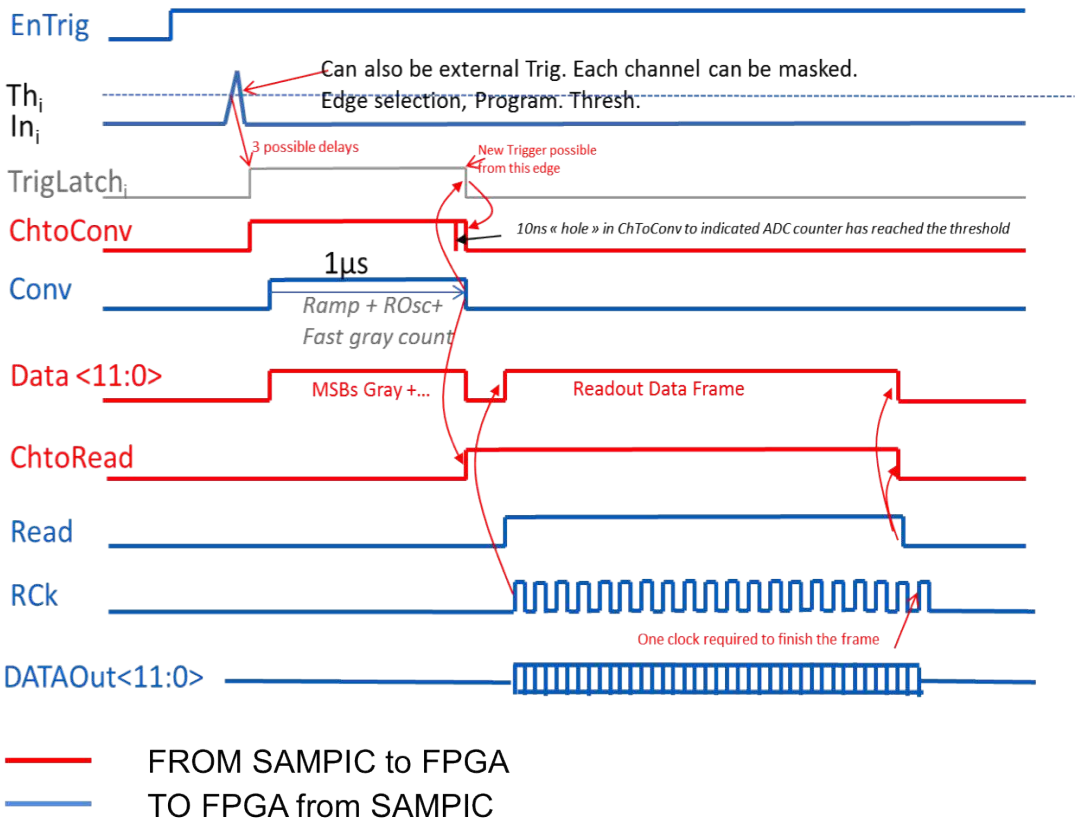


Figure 8.115 – Chronogram from the conversion and readout of one channel.

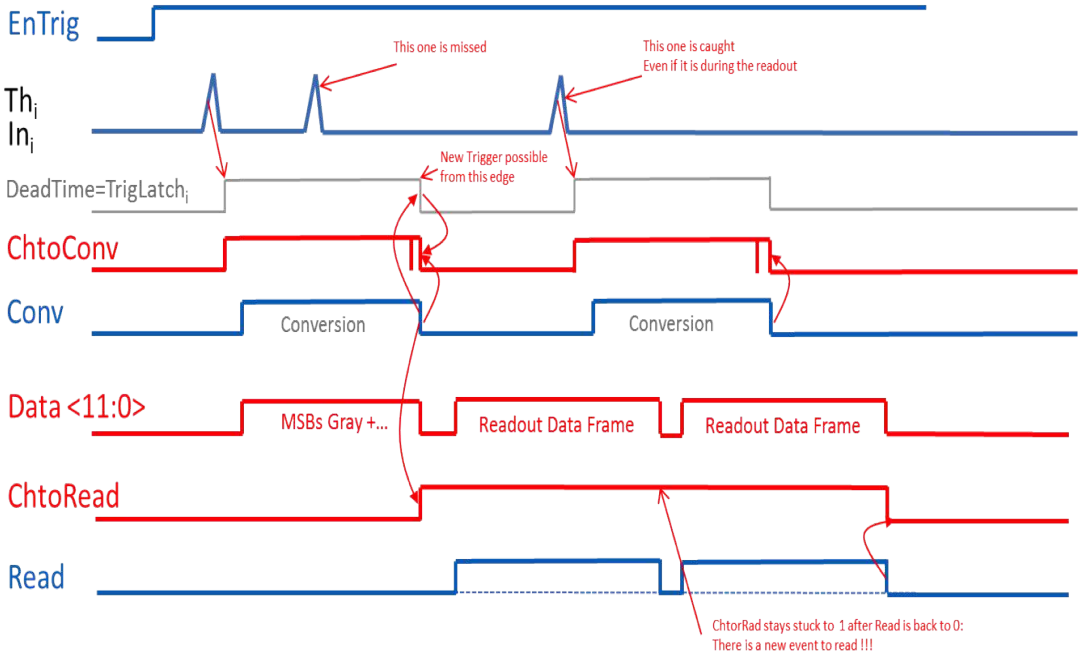


Figure 8.116 – Chronogram from the conversion and readout of one channel with multiple hits.

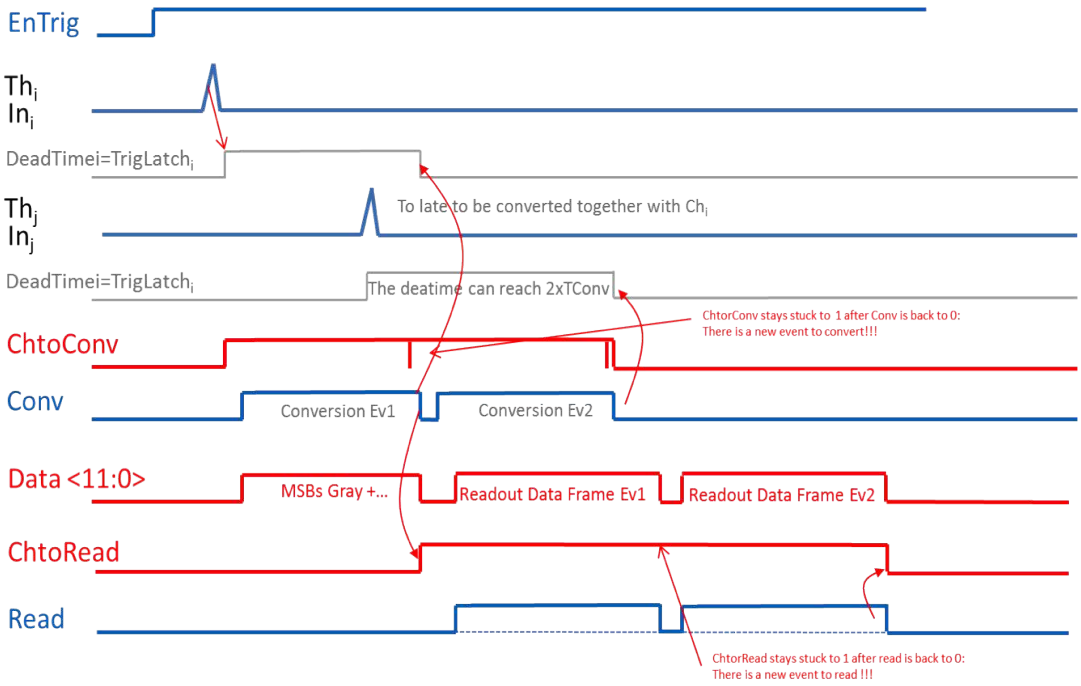


Figure 8.117 – Chronogram in the case of two hits in two channels converted separately and read sequentially.

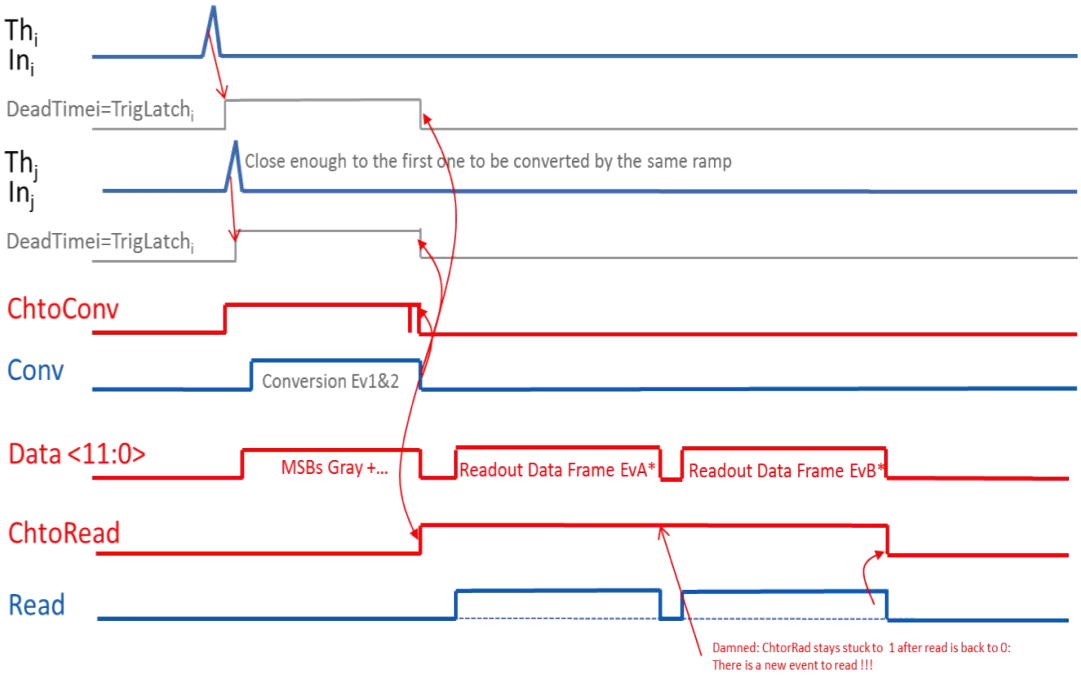


Figure 8.118 – Chronogram in the case of two hits in two channels, converted simultaneously and read sequentially.

8.13 Conclusion

The SamPic is a chip directly intended for measuring the time of a signal. During all steps of the designs special attention was paid to all parameters and features impacting the sensitivity of time measurement such as noise, jitter or mismatches. The chip was successfully submitted to the foundry and manufactured. Some results will be given in the following Chapter. A picture of the full chip layout can be seen in Figure 8.119

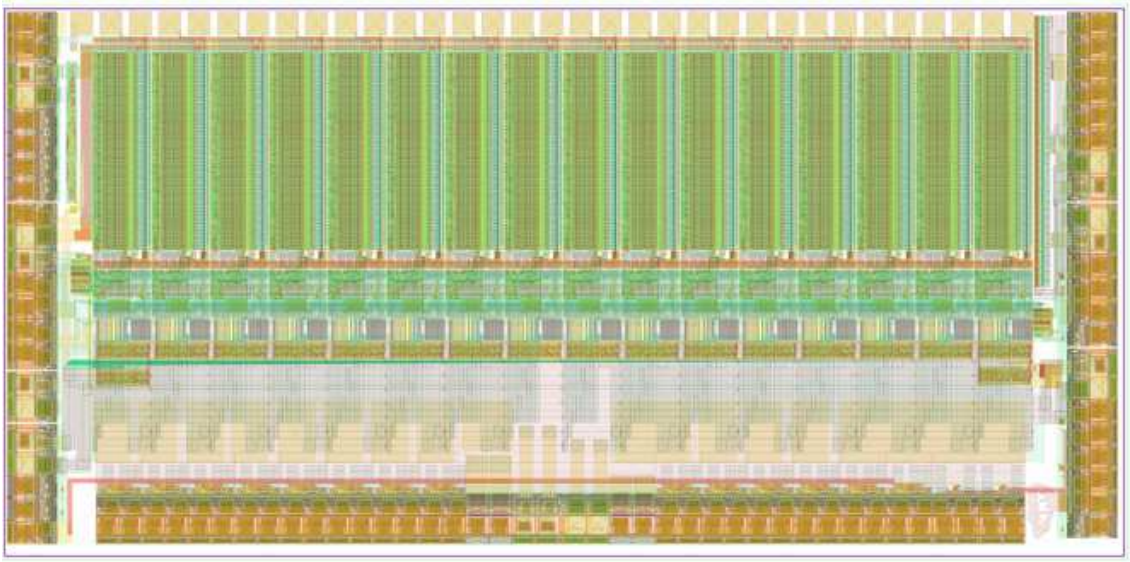


Figure 8.119 – Picture of the layout of the SamPic chip.

Bibliography

[1] Dominique Breton and Eric Delagnes. *Fast analog sampler with great memory depth*, February 22 2005. US Patent 6,859,375.

[2] *AMS*. www.ams.com.

[3] Henry Frisch. *The Factors that Limit Time Resolution in Photodetectors, Workshop, Univ. of Chicago, Chicago, IL.* http://psec.uchicago.edu/workshops/, April 2011.

[4] Razavi Behzad. *Design of analog CMOS integrated circuits*. 2001.

[5] Gary S Varner, LL Ruckman, JW Nam, RJ Nichol, J Cao, PW Gorham, and M Wilcox. *The large analog bandwidth recorder and digitizer with ordered readout (LABRADOR) ASIC*. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 583(2):447–460, 2007.

[6] Phillip E Allen and Douglas R Holberg. *CMOS analog circuit design*. Elsevier, 2011.

[7] Stefan Ritt. *The DRS chip: Cheap waveform digitizing in the GHz range*. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 518(1):470–471, 2004.

[8] Rudy J Van de Plassche. *CMOS integrated analog-to-digital and digital-to-analog converters*, volume 2. Kluwer Academic Publishers Dordrecht, 2003.

[9] Paul Horowitz, Winfield Hill, and Thomas C Hayes. *The art of electronics*, volume 2. Cambridge university press Cambridge, 1989.

[10] Stuart Kleinfelder, SukHwan Lim, Xinqiao Liu, and Abbas El Gamal. *A 10000 frames/s CMOS digital pixel sensor*. *Solid-State Circuits, IEEE Journal of*, 36(12):2049–2059, 2001.

[11] E Delagnes, Y Degerli, P Goret, P Nayman, F Toussenel, and P Vincent. *SAM: A new GHz sampling ASIC for the HESS-II front-end electronics*. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 567(1):21–26, 2006.

9 SamPic test results

Contents

9.1 Manufacturing and packaging

9.1.1 Die manufacturing

The SamPic chip was manufactured through CMP Grenoble, the French Multi-project Wafer Integrated Circuit Fabrication Services Provider. As described in Chapter 8, the design was made using the $0.18\mu m$ CMOS from Austria Microsystems. Forty chips where manufactured. A picture of the SamPic die is shown in Figure 9.1. Out of the forty chips, 10 of them were packaged inside a 128 pins ceramic Quad Flat Package (QFP)

The chip was almost fully functional at start-up [1] and is operating normally in all modes. Some preliminary results will be presented in the following Sections.

The configuration of the internal functions in SamPic are done via slow control that gives access in read and write mode to all the configuration registers. In SamPic, there are one configuration register in each channel for the triggering (internal threshold, triggering mode, ...) plus three global registers (two for the DLL and phase comparator configuration, one for global configuration).

[1] A connection was not made from the main power supply ring to the Power-On-Reset cell of the chip. Since this structure provides the reset signal to the readout state machine a connection had to be made and was done by focused ion beam by Presto Engineering in Caen.

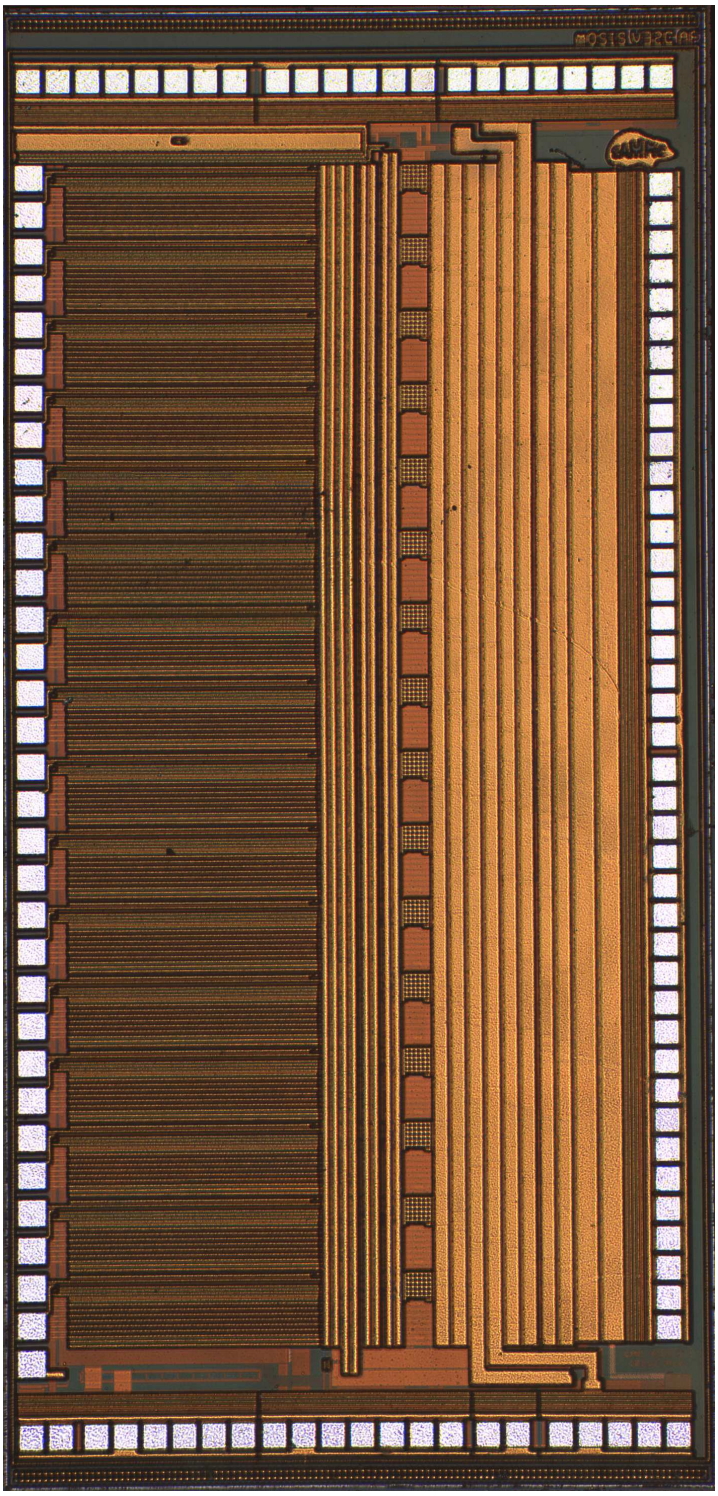


Figure 9.1 – Picture of the manufacture SamPic die. The size of the die is 1880μm by 3660μm.

9.1.2 Acquisition board

The test system for SamPic was developed at LAL Orsay and is shown in Figure 9.2. It is made of one mother board controlling separately up to two daughter boards, each one housing one SamPic chip. The chips are controlled by an Altera Cyclone III FPGA.

The boards are connected together with a vertical straight connector. They are powered by an external 5V Volt supply and the complete system power consumption is about 5W.

The connections to the input channels of the chip are made via 16 SMC connectors, and an external trigger signal can be supplied to motherboard using a dedicated SMA connector.

The board can be read by USB, Ethernet or Optical Fiber. Only USB was used for the tests reported here.

There are no external amplifiers at the input of the SamPic channels.

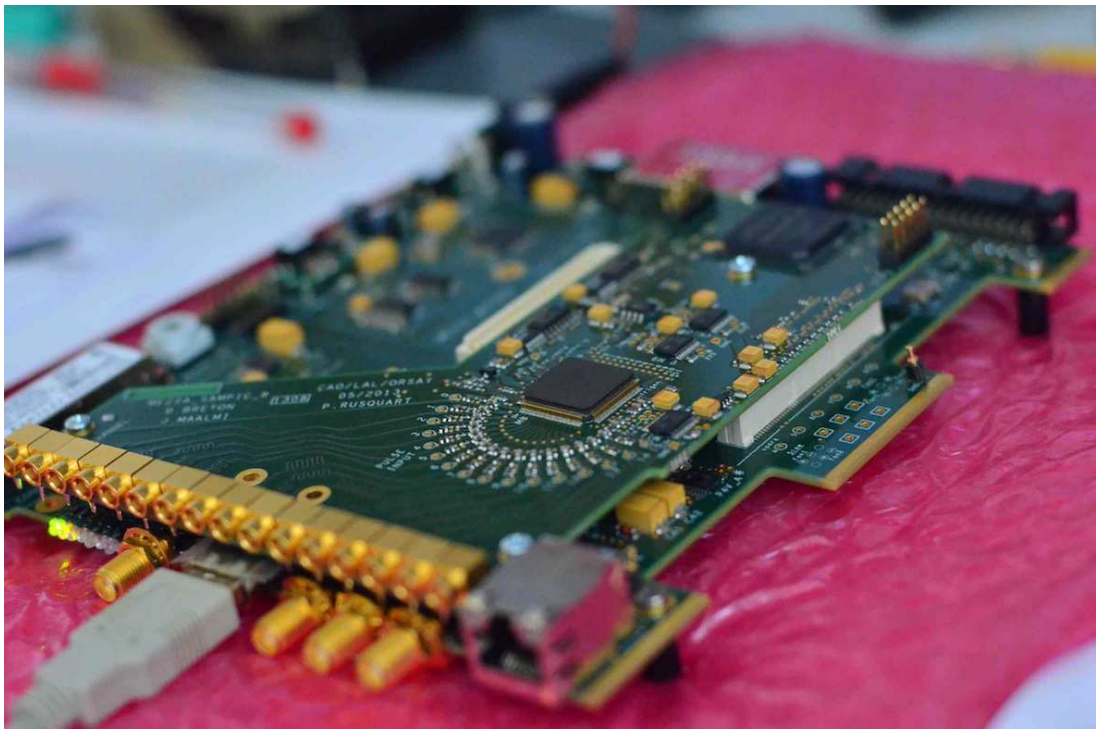


Figure 9.2 – Picture of the SamPic acquisition board. The connectors to the 16 channels are clearly visible, the line on PCB are matched in length to all inputs of SamPic.

9.1.3 Acquisition software

The acquisition software for SamPic was developed at LAL Orsay using the CVI framework. It allows to fully control and configure SamPic and it also integrates some measurements capabilities:

- Time delay measurement between channels using a CFD algorithm.

- Pedestal calibration.
- Linearity measurements.
- Noise measurements.
- Sampling time jitter measurements.
- Single event - Continuous - Finite Hit number - Finite time acquisition mode.
- Pulse charge measurement.
- Hit rate monitoring.
- Control of on-board internal pulsers for calibrations.
- Saving data to file.

The software is provided with a Graphical User Interface (GUI) displaying in real time the readout data - scope-like. It is shown in Figure 9.3.

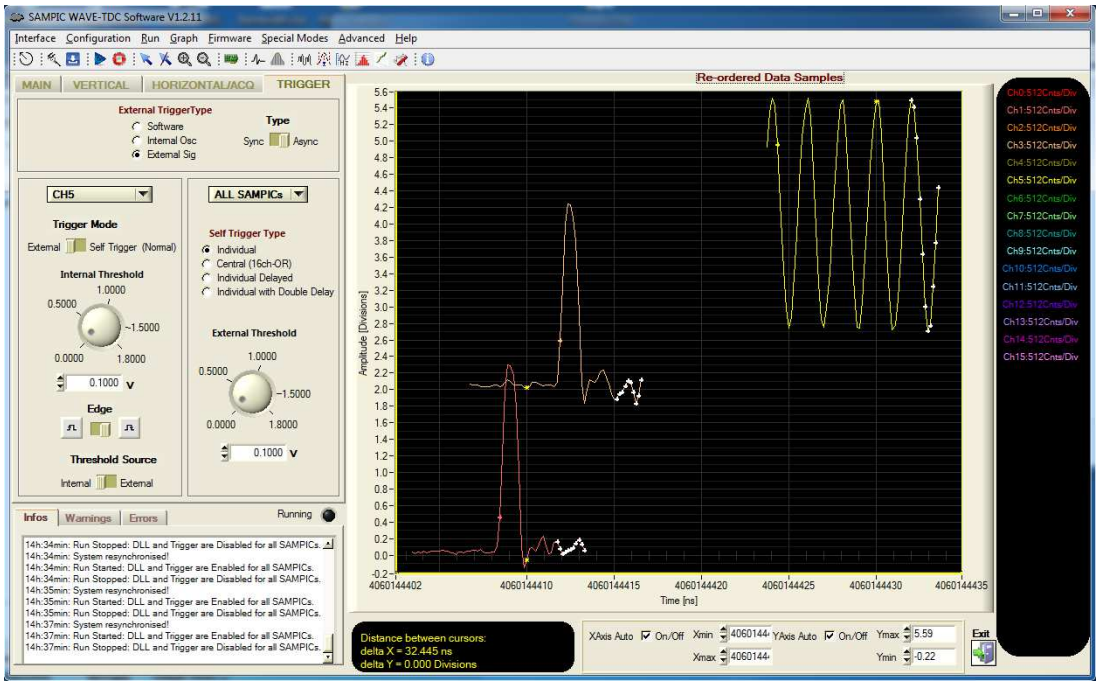


Figure 9.3 – Illustration of the GUI software used for the system acquisition and readout.

9.2 Chip performances

All 16 channels are fully functional up to 8.5 GSa.s^{-1}, the first 8 channels are functional at 10 GSa.s^{-1}. Actually, we have found that, at 10 GSa.s^{-1}, due to insufficient buffering, some sampling window do not propagate to the 8 last channels (see Section 9.5).

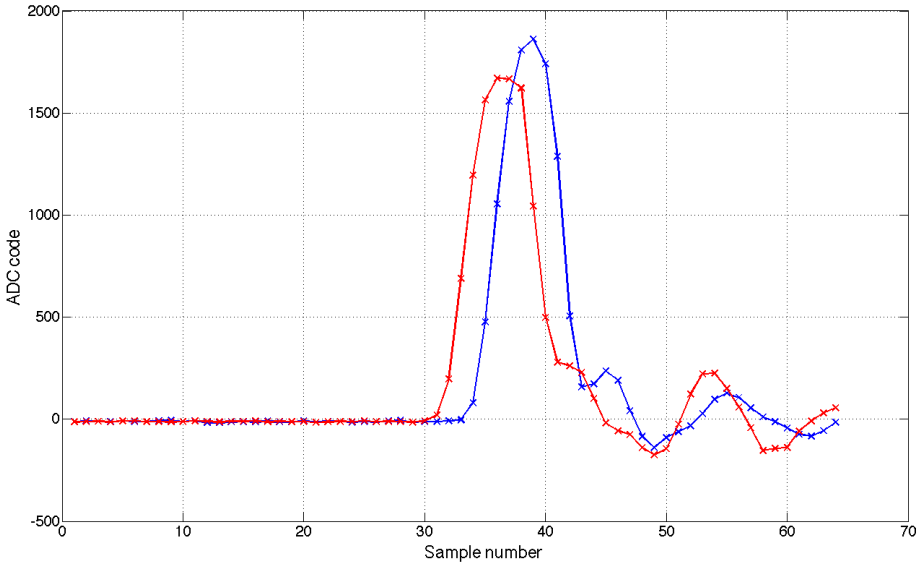


Figure 9.4 – Plot of two pulses, 1ns FWHM recorded simultaneously in two different channels of SamPic at 6.4Gs.s^{-1}.

9.3 ADC performances

All 1024 (16 by 64) Analog to Digital Wilkinson converters of SamPic are functional. The average RMS noise on all cells is measured to be 1 mV. The noise map of all cells is shown in Figure 9.5. The measured input dynamic range is 1 Volt on 11 bits.

The linearity of all the cells as also been measured and is shown in Figure 9.6, with a pedestal calibration done at mid-range. We observe a 2% peak-to-peak spread on the gain of all the cell, and a 3% peak-to-peak non-linearity. This non-linearity is mainly due to the non-linearity of the charge injection (see Figure 8.35) and of the comparator (see Figure 8.92). The non-linearity can be calibrated if necessary, but it has not been done in any of the results reported hereafter.

9.3.1 Trigger performances

Each of the 16 independent internal discriminators are functional. The S-curves of are shown in Figure 9.7 and Figure 9.8. A 2 mV RMS noise on the discriminator threshold is measured in the internal threshold mode (default mode).

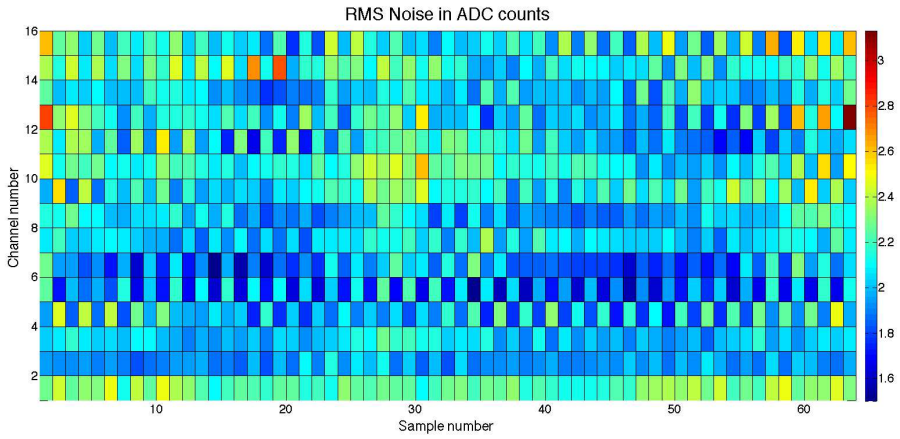


Figure 9.5 – Illustration of the RMS noise in the 16 × 64 ADC sampling cells of SamPic. We measure an average of 2ADC counts RMS corresponding to 1mV RMS noise.

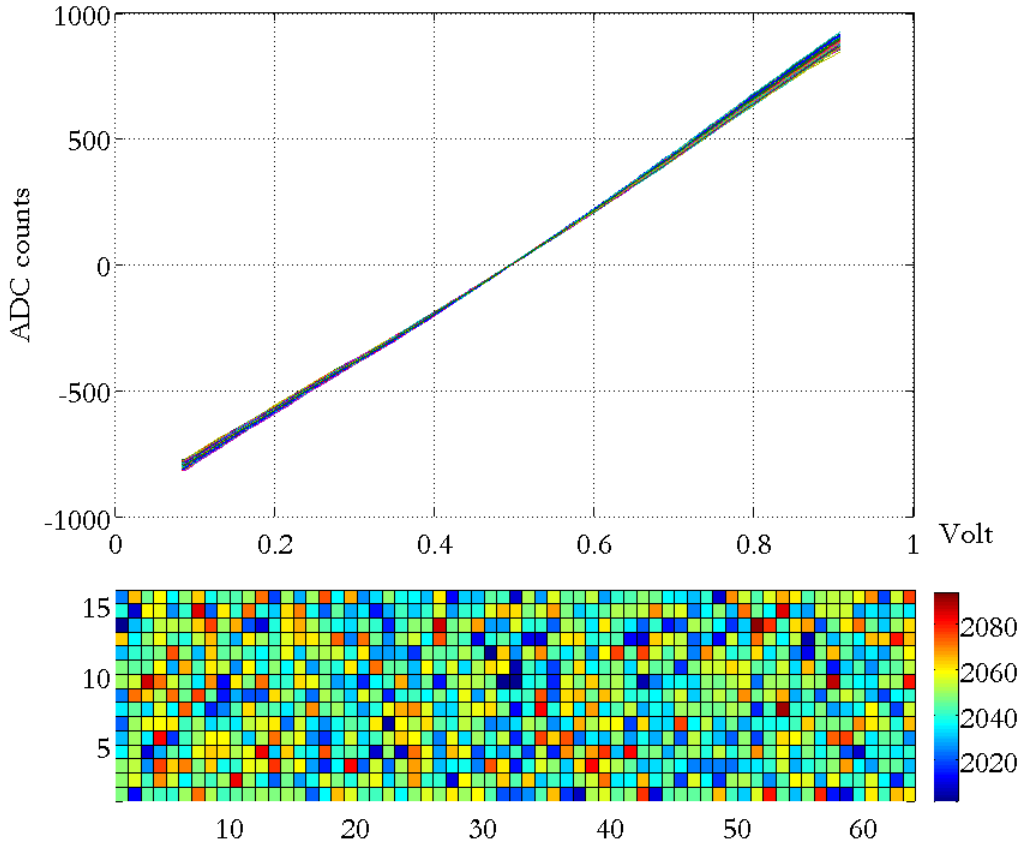


Figure 9.6 – Illustration of the linearity of the 16 × 64 ADC cells of SamPic. The pedestals have been calibrated at mid-range. The bottom noise map is the distribution of the gain in all cells.

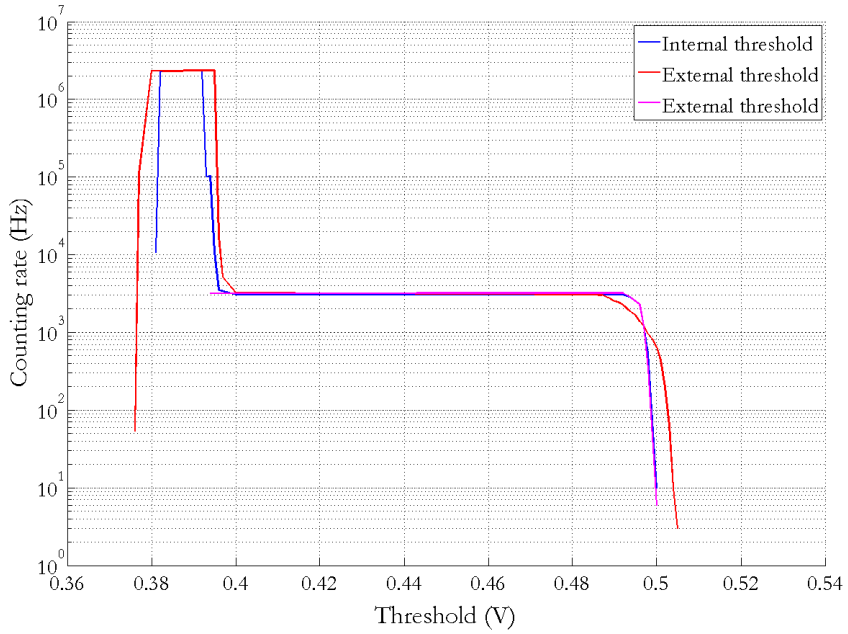


Figure 9.7 – Discriminator S-curve with 150 mV - 1ns FWHM pulse at a 3kHz rate. The S-curve for the internal threshold is in red and for the external (and common to 16 channel) threshold in blue.

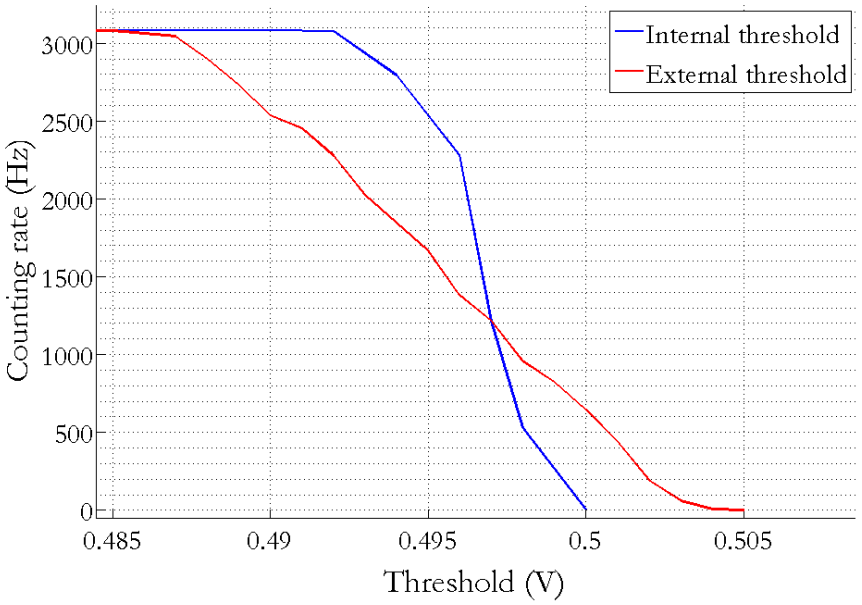


Figure 9.8 – Zoom on the right part of Figure 9.7 . At the maximum amplitude of the pulse a width of 8mV is measured corresponding to 2mV RMS noise for the internal threshold.

9.4 Timestamp

The time stamp records the state of the 12 bit coarse Gray counter of the SamPic chip in each channel when a trigger occurs. There is no missing code in the Gray counter. To resolve the metastability two value of the counter are latched: $Timestamp = TA$ and $Timestamp + ck_{ref}/2 = TB$. The hit distribution of $TB - TA$ is shown in Figure 9.9.

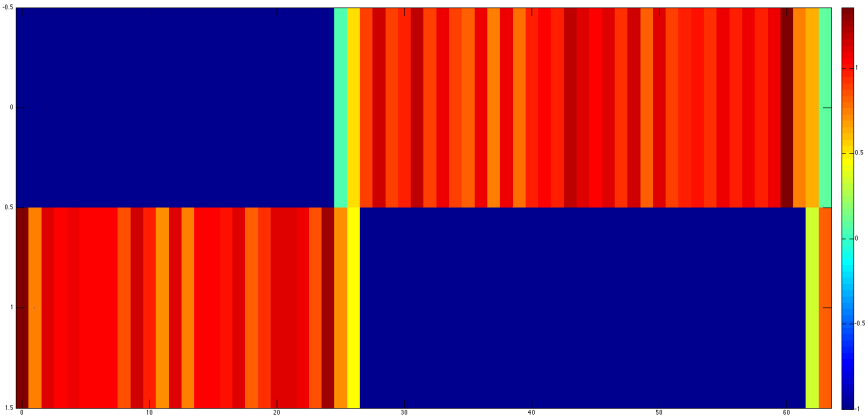


Figure 9.9 – Hitmap of $TB - TA$ along the 64 cells in SamPic, the sampling is done from left to right. In the first part (cells 1 to 25) $TA = TB + 1$, then $TA = TB$ (from cells 28 to 62) the metastability zones are cells 27, 28 and cells 63, 64.

In the first part (cells 1 to 25) $TA = TB + 1$ with a 100% probability. From cell 28 to 62, $TA = TB$ also with a 100% probability. The cells 27, 28 and 63, 64 where the two possibilities are co-existing correspond to metastability regions. To avoid them, the value chose for the timestamp is:

- TA for cells in the $[2, 58]$ range.
- $TB + 1$ for cells in $[0, 1]$ and $[59, 63]$.

9.5 Sampling window

SamPic employs a specifically design constant size sampling window equal to 11 $\times t_{step}$. The histogram of the sampling window sizes for 200 000 events, in Figure 9.10, shows that the sampling window is always 11 or 10 cell wide. This is true up to 8.5 GSa.s^{-1} for all the channels. At 10 GSa.s^{-1}, only the first 8 channels receive correctly the sampling window from the DLL.

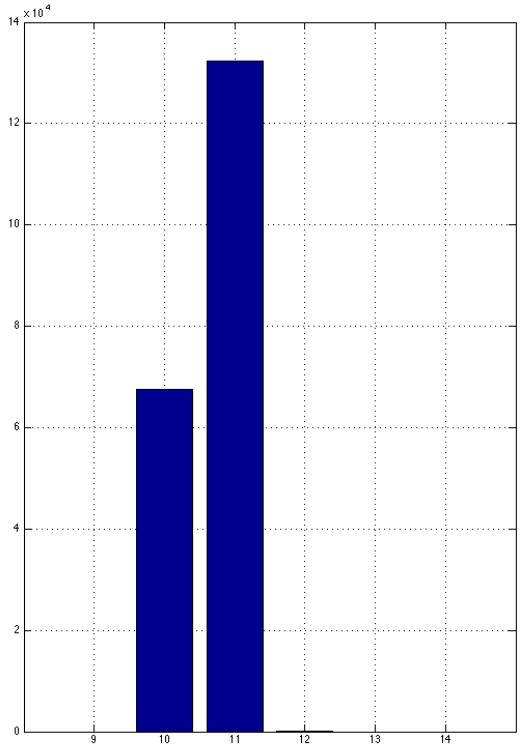


Figure 9.10 – Histogram of the sampling window size in SamPic done for 200 000 events.

9.6 Leakage

Leakage was one major concern of this technology. A plot of the difference between the sampled value, converted immediately after trigger and with a delay of 700 μs, is shown in Figure 9.11. We observe the exponential dependence of the leakage current predicted by simulations (see Figure 8.23b) close to the rails. However, the linear dependence is not observed. For a stored voltage $V_{store} = 0$ V, 5 pA of leakage current can be extrapolated from this measurement. For $V_{store} \geq 100mV$, this current dropped dow to below 5 fA and corresponds to negligible droop for tens of micro-seconds storage time. This value is better than the 10 fA expected from our simulations.

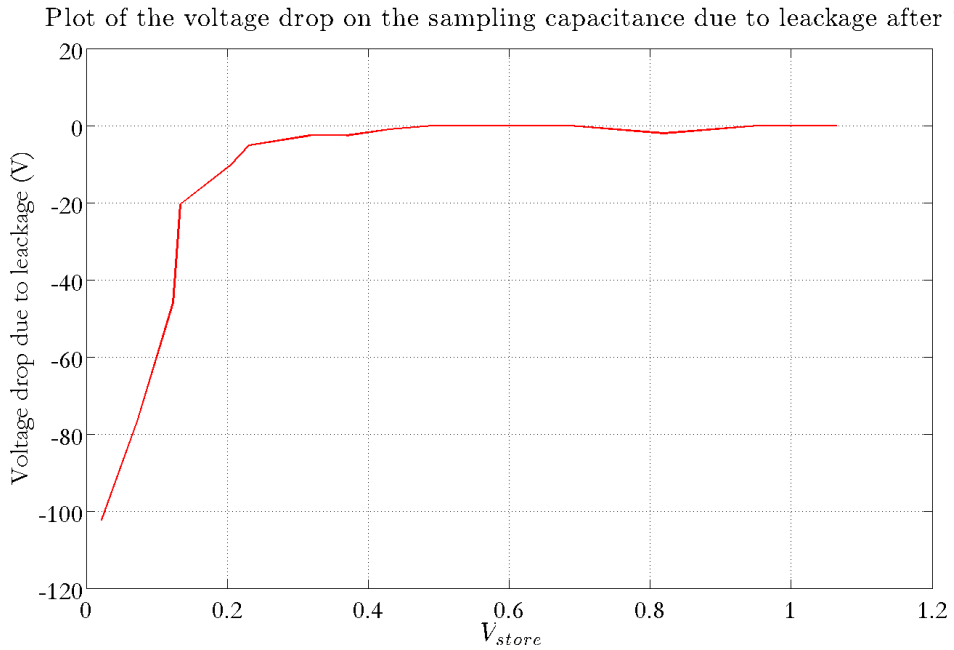


Figure 9.11 – Plot of the difference between the sampled value converted immediately after trigger and with a delay of 700 μs in SamPic as a function of the stored voltage $V_{store} = 0$ V. The largest value corresponds to 5 pA of leakage current.

9.7 Bandwidth

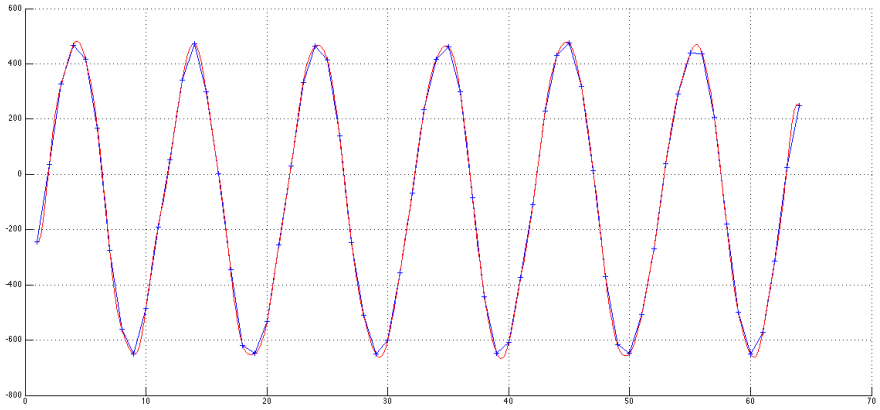


Figure 9.12 – 1GHz sinewave sampled at 10Gs.s^{-1} in the first channel of SamPic. The blue dots are the 64 data samples and the red curve is a *sinc* interpolation on the points.

The SamPic frequency response directly affects its timing capabilities. In order to appreciate it, a sinewave with a 500 mV peak-to-peak amplitude is sent to one channel (channel 2). The gain measured on this channel and its neighbors (for cross-talk extraction) is plotted in Figure 9.13 versus the sinewave frequency. A 1.6 GHz -3dB bandwidth is measured, and is similar for all the channels. Moreover, this value is constant over the 64 cells of each channel, thus validating the uniformity of the input line layout. The magnitude and the shape of the cross-talk showing ringing at 1.2 GHz is still under investigation.

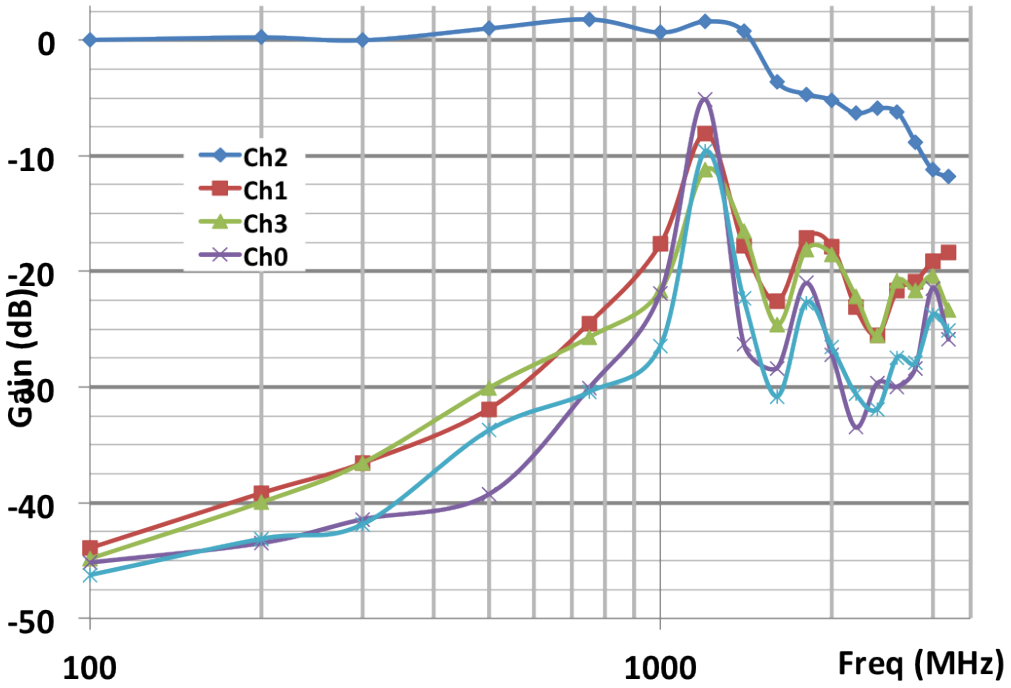


Figure 9.13 – Measurement of the SamPic bandwidth on channel 2 and crosstalk on adjacent channels. The 3dB bandwidth is 1.6 GHz. The reason for the resonance peak at 1.2 GHz is unknown and being investigated.

9.8 Time resolution

A calibration of the fixed pattern offsets in the sampling times $t_i \quad i \in [0, 63]$ is required for precise time measurements [1]. This principle is illustrated on the Figure 9.14. The measurement of the Differential Non Linearity (DNL) and of the Integral Non Linearity (INL) is shown in Figure 9.16.

The INL gives a good idea of the resolution of time measurements without corrections and of the possible improvement with INL corrections.

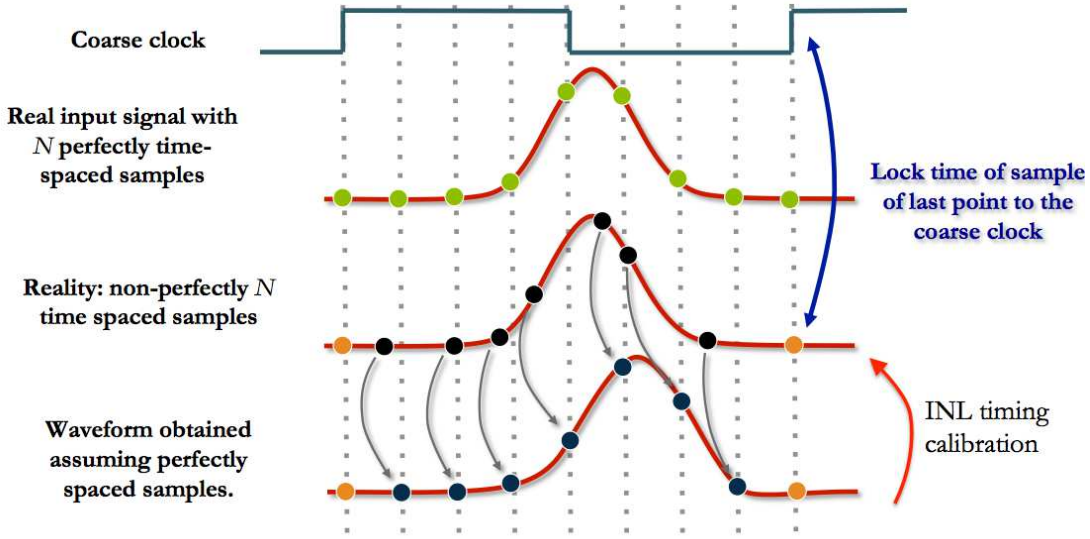


Figure 9.14 – Illustration of the principle of Integral Non Linearity (INL) correction for the sampling times $t_i \quad i \in [0, 63]$.

Excellent time resolution has been measured with SamPic, we report here a RMS precision between **5 and 10ps** on the time delay estimation between two pulse delayed from 0 to 65ns - the pulses were delayed using cables which might be responsible for the degradation of the time resolution for higher delays (higher cable length). These results are shown in Figure 9.17, the algorithm used was CFD. The distribution of the delays for $\Delta t = 2ns$ is shown in Figure 9.18.

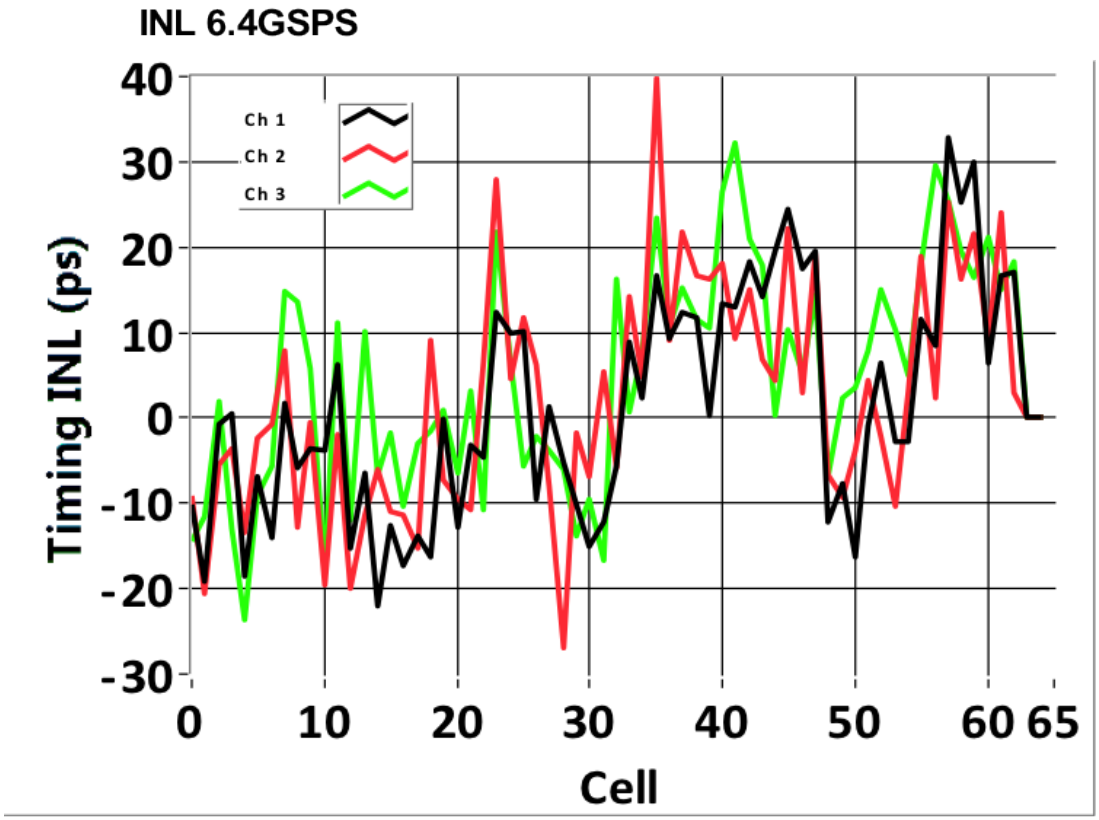


Figure 9.15 – Plot of the Integral Non Linearity (DNL) of the three first channel of SamPic at 6.4Gs.s^{-1}. In the INL shapes of all channel we can see a common pattern which is probably characteristic from the DLL and second order patterns probably due to the in-channel logic.

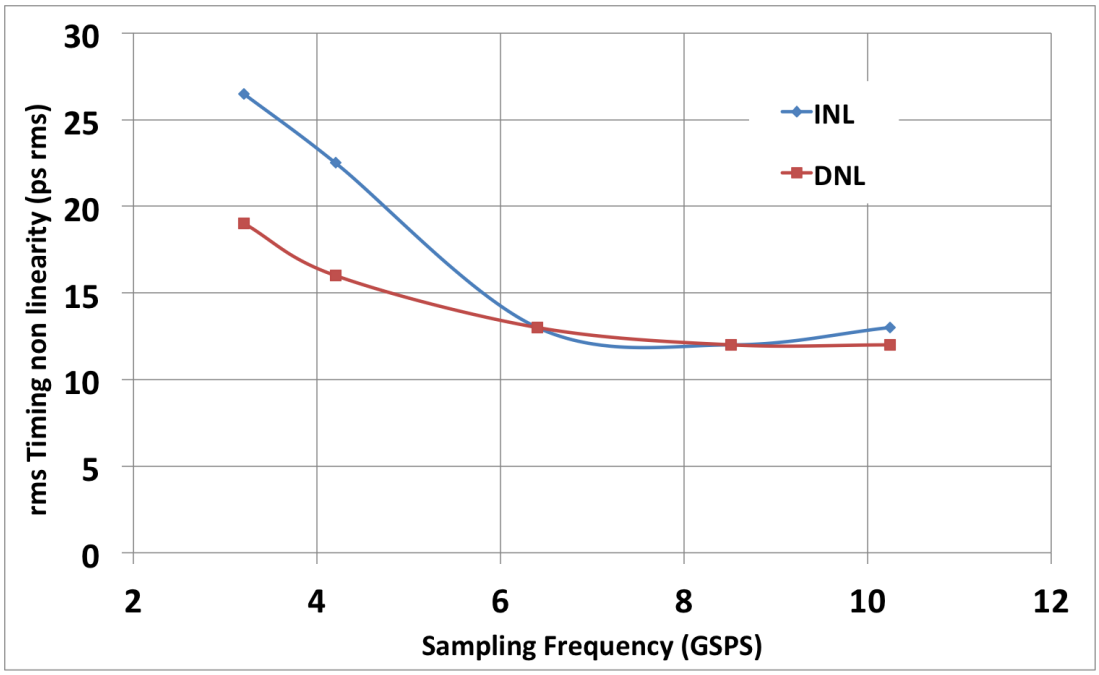


Figure 9.16 – Plot of the Differential Non Linearity (DNL) and of the Integral Non Linearity (INL) at 6.4Gs.s^{-1} for channels 1, 2 and 3.

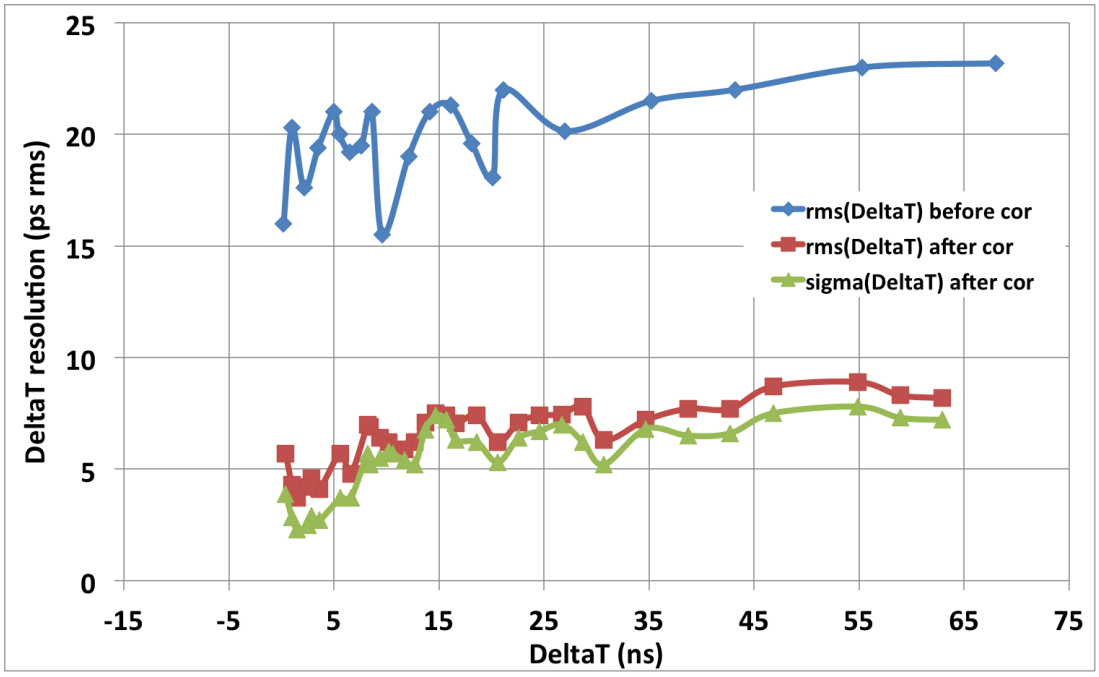


Figure 9.17 – Measurement of the time delay estimation between two pulse delayed from 0 to 65ns. We report here a sub-10ps RMS precision across the entire range.

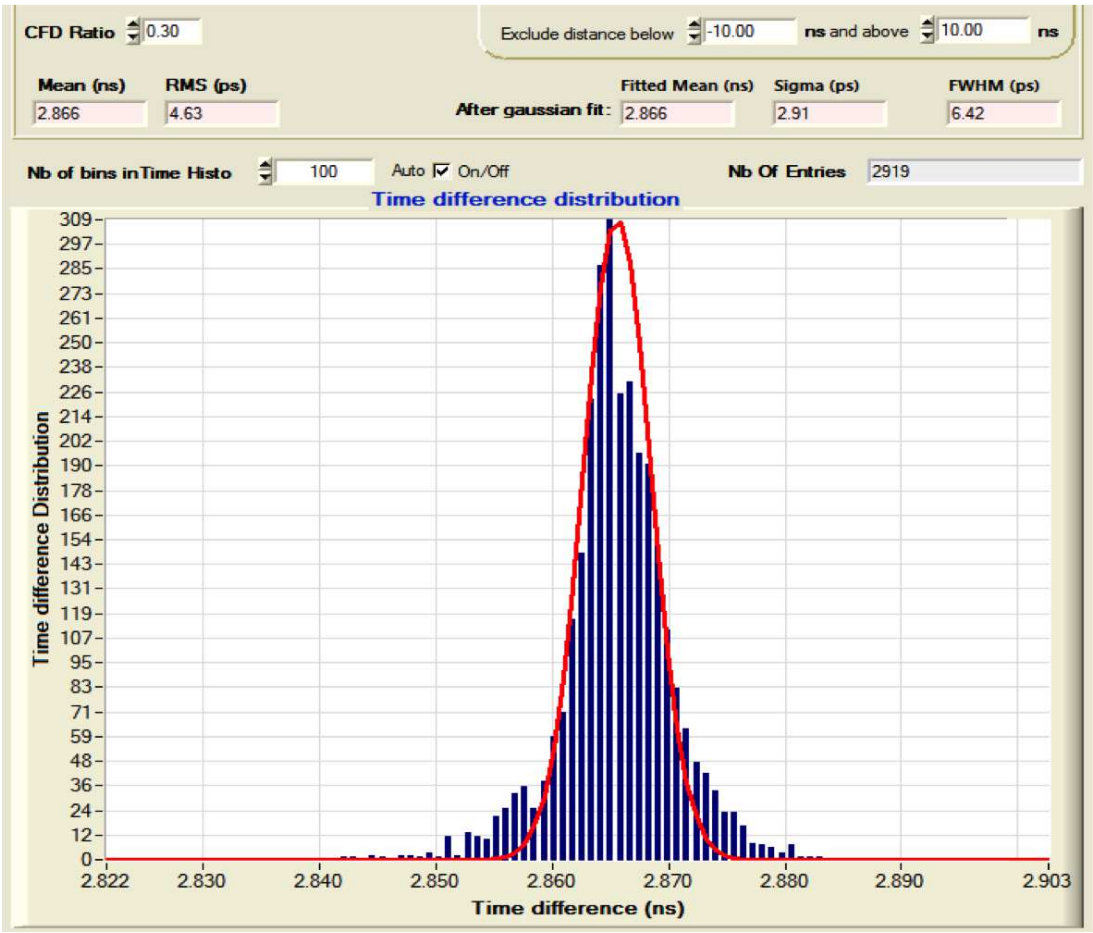


Figure 9.18 – Histogram and fit of the delays measured with SamPic for two pulses spaced by 2.8ns. Here the RMS noise of the measurement on the individual pulse is 4.6ps.

9.9 Power consumption

The SamPic chip power consumption is 10mW per channel for a sampling seep of 6.4Gs.s^{-1} in low current mode for the LVDS. The total power consumption for each functionality in SamPic is shown in Figure 9.19. During the design the power supplies for each function (ADC, Sampling, Readout, Discriminator, Slow Control, DLL, ...) have been made independent (separate input pads).

- The Delay line and its output buffers to the 16 channels is one of the main power consumption in the chip, totalling 50mW
- The Wilkinson Analog to Digital Converters use $2\mu W$ per cell.
- In the high current mode the LVDS output drivers use 70mW, which scales down to 7mW in the low power mode (current divided by 10).
- The other main contributor of the power consumption in SamPic is the trigger and sampling logic (see Section 8.10 of Chapter 8), responsible of 60mW of power consumption.
- The internal discriminators use 1.1mW per channel.

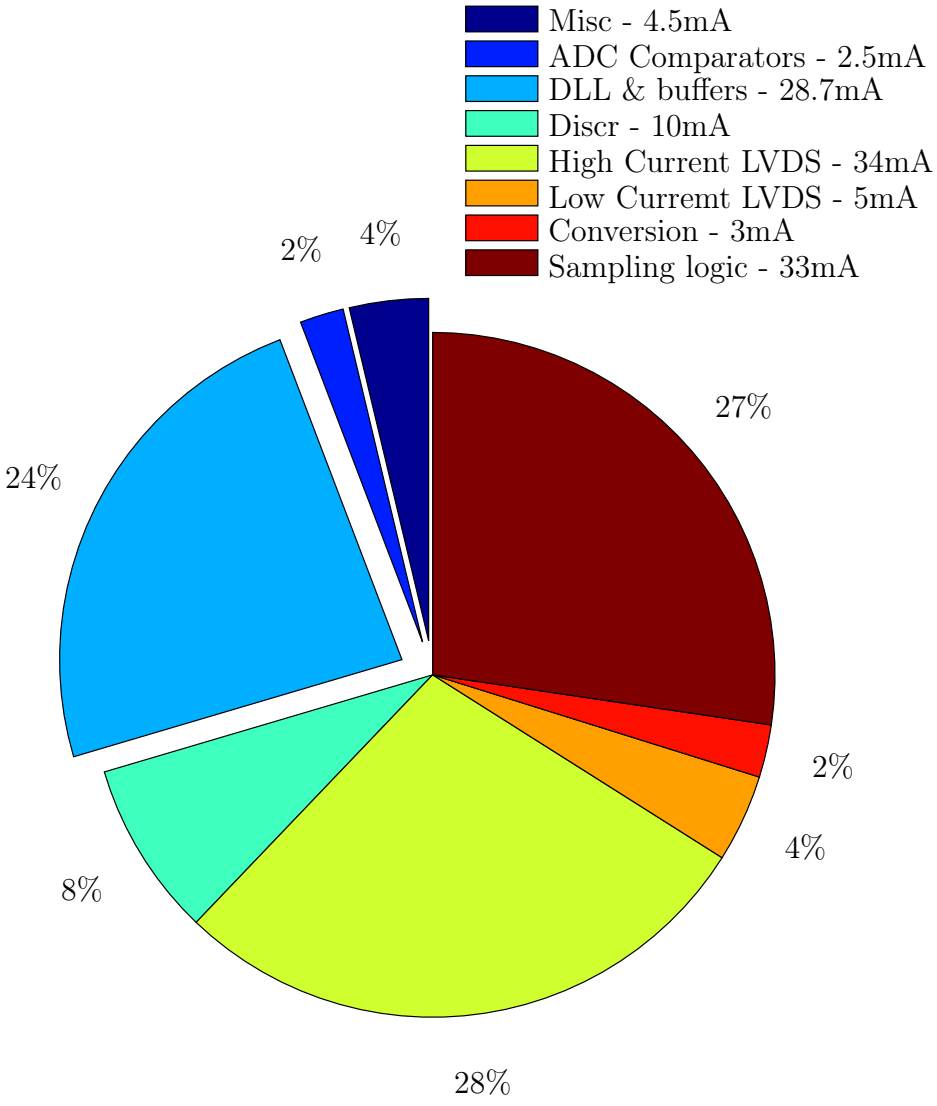


Figure 9.19 – Report on SamPic power consumption.

9.10 Conclusion

The preliminary results shown here are very encouraging especially the precision achieve on the time measurements. A table summarizes the SamPic chip main performances.

Table 9.1 – Table of SamPic main performances.

	Performances measured
Number of channels	16
3dB Input bandwidth	1.6 GHz
Sampling frequency	3.2 - 10 GSa.s^{-1} [2]
Number of data samples	64
ADC precision	11 bits
Noise	1 mV RMS
Input dynamic range	1 V
Readout clock	160 MHz [3]
Conversion time	1.6 μs for 11 bits - 200 ns for 8bits
Readout speed	1.92 Gb.s^{-1}

[2] Only 8 first channel functional at 10 GSa.s^{-1}

[3] 400Mhz possible but not verified

Bibliography

[1] D Breton, E Delagnes, J Maalmi, et al. *Picosecond time measurement using ultra fast analog memories*. In *Proceedings of the Topical Workshop on Electronics for Particle Physics-CERN-2009-006*, pages 149–154, 2009.

Conclusion

In this thesis dissertation, we have presented the conception of a picosecond time-of-flight measurement system for the ATLAS experiment. This research work was based both on the Large Area Picosecond Photo-Detector, at the University of Chicago, and the SamPic, at the CEA, projects.

The study of the sensitivity to Beyond Standard Model physics, brought by the design and installation of picosecond time-of-flight detectors in the forward region of the ATLAS experiment at the LHC and presented in the first part of the dissertation shows that, using the AFP detector at LHC, will bring a significant improvement to the measurement of quartic gauge anomalous couplings between the photon and the W boson, using exclusive WW pair production in ATLAS. Here, the semi-leptonic decay considered for the WW pair, allows to probe for anomalous couplings a_0^W in the order of 5.10^{-6}, similar to the pure leptonic case, and thus bringing additional sensitivity in the search for quartic gauge anomalous couplings at LHC.

In order to be able to make this measurement, we require the installation of a precise timing and tracking detector in ATLAS: the AFP detector. The AFP detector deployment in the LHC tunnel is presented in the first part of the dissertation. In the second and third part, we present a description of large area picosecond photo-detectors conception, time reconstruction algorithms and custom readout integrated circuit with a special care given to signal sampling and processing for precision timing.

The description of large area picosecond photo-detectors conception is based upon the research and development work realized at the University of Chicago, within the LAPPD collaboration. A special attention has been given to all the detector parts critical for precise timing. As of today the 20 by 20 cm detector is very well advanced with all active parts: photo-cathode, micro-channel plates, anode array and electronic readout successfully designed. Pulses with rise-times below 1 ns have been observed when shining a laser source on the detector with external vacuum applied.

In order to cope with the high rate and low dead-time required for the AFP detector at the LHC experiment a dedicated Application Specific Integrated Circuit (ASIC) called SamPic has been designed to provide data readout of the timing detector. The ASIC has been successfully designed, manufactured and tested. It presents the best precision in this domain with an impressive timing resolution of 5 to 7 picosecond for signals delayed from 0 to 64 ns.

The field of precise timing is growing in popularity not only in the particle physics world but more generally in the imaging field; a precision of 10 picosecond in time-of-flight for light being equivalent to a 3 mm depth resolution, the industry is more and more attracted by this new technology in order to get higher resolution medical scans using PET Time-Of-Flight, more precise rangefinders or night vision goggles in the defense industry or 3-D imaging and positioning of players in the gaming industry.

10
Résumé en Français

Contents

10.1 Introduction

Depuis 2009, date à laquelle j'ai commencé a travailler sur le sujet de la mesure de temps picoseconde à l'Université de Chicago, la physique expérimentale à largement profité des innovations de l'industrie micro-électronique qui permet la conception de circuit plus rapide, moins cher et moins energivore. La précision de la mesure du temps, un temps limité par les limitations de l'échantillonnage et du traitement du signal provenant des détecteurs de particules, a gagné un ordre de grandeur en vitesse et précision grace à l'utilisation de mémoire analogique ultra-rapide. Ceci a permis d'atteindre des précisions de mesure sur les mesure de temps de l'ordre de la picoseconde.

Simultanément, les améliorations des processus de fabrications et les diminution des coûts de frabications des photo-détecteurs rapides, conséquence de la démocratisation de l'utilisation des capteurs d'images dans les équipements numérique d'aujourd'hui; mais aussi le fruit d'une longue histoire de dévelopement de détecteur pour l'imagerie (détecteur de pixel au CERN ou bien imagerie médicale) et d'amélioration de la robustesse des photo-multiplieurs maintenant largement disponible dans la communauté de la physique des hautes énergie. Des programmes de recherche et dévelopement comme le project *Large Area Picosecond Photo-Detector* contribuent aussi à l'amélioration de la résolution et de la taille des photo-detecteurs actuels, tout en minimisant leur coût de production.

Cette combinaison d'améliorations permet aujourd'hui de nouvelles possibilités de mesure dans la physique des particules.

À la vitesse de la lumière, une précision de mesure d'une picoseconde dans la mesure du temps de temps de vol correspond à une précision de 0.3 mm en distance parcourue. Cette précision inégalée de mesure permet, en physique des particules, d'accéder à de nouvelles classes de mesures. En effet, les efforts pour observer des particules prédites dans le cadre du Modèle Standard, comme le boson de Higgs (observé le 4 Juillet 2012) ou découvrir une nouvelle physique au-delà du Modèle Standard, ont conduit à augmenter le nombre de particules interagissant simulatanément dans les collisionneurs de particules, produisant jusqu'à 100 interactions simulatnée au LHC (comparativement aux 6 du Tevatron). Ce nombre important d'évènement simultanés impose de nouvelles méthodes afin de pouvoir identifier les vertex créé. Une de ces méthode est la mesure du temps de vol, ou les particules créé sont identifiée à leur vertex d'origine en utilisant le mesure de leur temps de parcours depuis leur origine jusqu'à leur point de détection. Une résolution d'une picoseconde, correspondant à une résolution de 0.3 mm dans la position du vertex, conduit à une segmentation suffisante en comparaison aux 20 cm de la taille de la zone d'interaction.

Dans cette thèse, nous étudions, dans une première partie, l'installation ainsi que l'intérêt de détecteur de timing rapide dans l'expérience ATLAS au CERN. Nous étudierons en particulier les améliorations possible dans la sensibilité de la mesure au couplage anormaux entre photons et le boson W. Dans la seconde partie sont décrit les améliorations nécessaires pour la fabrication de photo-detecteur rapide ainsi que le traitement du signal requis pour une mesure précise du temps.

Enfin dans la troisième partie est présenté le dévelopement d'une micro-électronique spécifique pour la mesure précise du temps (l'ASIC SamPic) et les performances obtenues par l'ASIC.

10.2 Physique des hautes énergies

Dans cette partie est présenté le cadre de Physique des Hautes Energies dans lequel cette thèse à été réalisé. Dans une première partie est décrit le Large Hadron Collider.

10.2.1 Le LHC

Le Grand collisionneur de hadrons (LHC) installé dans un tunnel à la frontière de la France et de la Suisse, est le plus grand accélérateur de particules du monde. Il a été ignauguré le 10 Septembre 2008, après dix ans de travaux d'installation. Il a une circonférence de 27 kilomètres, divisé en huit secteurs indépendants. Dans le tunnel, 1232 aimants dipolaires supraconducteurs sont utilisés pour courber deux faisceaux de particules de haute énergie. Les faisceaux voyagent dans des directions opposées dans des tubes séparés sous ultravide. Au LHC, les collisions principale sont des collision de proton, mais il est également conçu pour fonctionner avec des ions lourds.

Le LHC n'est pas un cercle parfait et est composé de huit arcs ainsi que de huit régions d'insertion. Les arcs sont destinés à courber les faisceaux et peuvent contenir jusqu'à 154 aimants dipolaires chacun. Les régions d'insertions sont des sections droites qui relient les arcs. La connexion entre eux est effectuée dans une région de transition dite région de suppression de dispersion, qui permet de réduire la dispersion de la machine (compensations des offsets) lors des insertions. L'architecture de la section droite dépend de son utilisation spécifique : physique (collision de faisceau au sein d'une expérience), l'injection, le dumping faisceau ou le nettoyage de faisceau.

La fréquence moyenne de collision f est différente de l'espcement temporel entre les bunchs (25ns correspondant à $40MHz$), en effet tous les emplacements disponible dans le faisceau ne sont pas remplis. Le délai moyen entre deux collisions est donc plus élevé (31.7ns) et la fréquence moyenne de collision est donc $f = 31.5MHz$. En raison de la décroissance de la luminosité, le nombre moyen d'interactions par bunch crossing μ diminue dans le temps.

Pour une luminosité initiale $\mathcal{L} = 10^{34}cm^{-2}s^{-2}$ et une section efficate totale $\sigma_{tot} = 100mb = 10^{-25}cm^2$, on obtient $\mu \approx 32$ interactions par bunch crossing. Pour luminosité supérieure, ce nombre augmentera à des valeurs comprises entre $\mu = 50$ à $\mu = 100$.

10.2.2 ATLAS

Le rôle des quatre expériences principales installés au LHC - ATLAS, CMS, ALICE et LHCb - est de détecter et d'identifier les particules secondaires produites lors des collisions aux quatre points d'interaction du LHC afin de reconstruire et d'analyser la physique de chaque événement. Pour ce faire, l'objectif du detecteur est de mesurer la position, la charge, la vitesse, la masse et l'énergie de toutes les particules créées. Par conséquent, les détecteurs de particules doit être hermétique, de sorte que seule une faible fraction des particules puisse s'échapper sands être détectés. Par commodité d'ingénierie, les détecteurs ont adopté la géométrie dite "baril plus bouchon", conception où un détecteur cylindrique couvre la région centrale et deux "bouchons" plats circulaires couvrent les angles proches du faisceau (la région avant).

Dans les différentes Sections sont présenté les concepts de base de la détection de particules dans les expériences de physique dea hautes énergies ainsi que le détecteur ATLAS dans plus de détails.

Une vue en coupe d'ATLAS (A Toroidal LHC) détecteur est illustré à la Figure 2.2. Elle révéle sa structure interne. De la partie la plus profonde de l'extérieur, il est composé du système de traces interne, les calorimètres électroniques et hadroniques et enfin les détecteurs à muons.

Le but du trajectographe interne est d'identifier les vertexes d'origine des particules chargées. Il est entouré par un solénoïde supraconducteur mince qui génére un champ magnétique uniforme de 2 Tesla. Le calorimètre d'ATLAS utilise l'argon liquide comme milieu actif à la fois pour les calorimètres électromagnétique et hadronique. Il est à noter que l'échantillonnage du signal se fait en utilisant des capacité commutées (voir l'architecture de SamPic au chapitre 8). Le système à muons d'ATLAS est placé des deux côtés d'unaimant toroïdal qui génére 0,5 Tesla en son centre.

Globalement, le détecteur ATLAS est un cylindre d'une longueur totale de 42 mètres et de 11 mètres de rayon.

Le détecteur ATLAS est un détecteur générique, qui permet une multitude de mesures de physique. La liste suivante résume les capacités de détection globales du détecteur ATLAS:

- La reconstruction de particules chargées pistes est possible jusqu'à $|\eta| \leq 2,5$.
- détection de muons peut être effectuée dans une gamme un peu plus grande $|\eta| \leq 2,7$.
- La plus grande couverture est assurée par les calorimètres $|\eta| \leq 4,9$.

10.2.3 Le détecteur AFP

La proposition de detecteur ATLAS Forward Proton (AFP) a l'intention d'explorer des événements pour lesquels les protons sont intacts dans l'état final [1] dans le détecteur ATLAS. Ceci permet l'étude d'une nouvelle gamme de processus de physique - typiquement des évènement élastique ou bien de diffraction - dans lesquels le proton émet un objet virtuel incolore appelé poméron, ou un photon, qui sera étudié ici. Les protons restent intacts après l'interaction et ont une impulsion plus faible par rapport au faisceau. Les aimants du LHC placé le long du faisceau sont utilisés comme un spectromètre en courbant la trajectoire de ces protons ce qui permet leur détection dans les détecteurs dédiés situé à proximité du faisceau, loin du point d'interaction d'ATLAS. L'idée du projet AFP est d'installer de tels détecteurs à environ 210 m du point d'interaction. Le suivi des protons à travers les aimants du LHC et et la mesure de leur impulsion permet l'analyse de la nature de l'interaction. En outre, le détecteur AFP sera entièrement intégré avec le détecteur principal, de sorte que les produits de désintégration de l'objet produit par diffraction pourront être mesurée dans le détecteur ATLAS. Cette physique n'a jamais été étudié auparavant avec de telles énergies de centre de masse. Les mesures effectuées avec AFP serait donc unique et conduire potentiellement à de nouvelles découvertes en physique.

[1] Événements inélastiques seulement car dans AFP l'acceptance des événements élastiques est négligeable

Dans la section 3.2 nous présentons le détecteur AFP et ses objectifs. Dans la section 3.3 nous présentons l'installation de détecteurs au LHC. Dans la section 3.4 nous présentons le détecteur de traces d'AFP et dans la section ref sec : timingdetctd le détecteur de timing.

Le but du détecteur de trace est de fournir une mesure des points le long de la trajectoire du proton et donc la mesure de la valeur du ξ des protons diffractés. La zone à couvrir par le tracker est approximativement de 20 par 20 mm. Pour ce faire, il a été décidé d'utiliser des détecteurs à pixels en silicium instrumentés avec le circuit intégré de lecture FE-I4, conçu à l'origine pour le Insertable B-Layer (IBL), une mise à niveau du détecteur de pixel ATLAS.

Le détecteur de timing d'AFP est nécessaire pour supprimer les événements de pile-up (rejeter les protons provenant d'événements qui se chevauchent). Afin de conserver les informations temporelle a une précision de quelques picosecondes, la détection du signal utilise le rayonnement Cerenkov créé par les protons dans une succession de barre de quartz.

10.2.4 La recherche de couplage anormaux entre les photon et le boson W dans ATLAS au LHC

Dans ce chapitre, nous présentons la perspective de la sensibilité aux couplages anormaux quartic entre photons et les bosons W, qui a été réalisé pour l'expérience ATLAS dans le cadre de l'installation des détecteurs AFP dans les régions avant. Le but de cette étude est d'investiguer le potentiel de découverte de nouvelle physique lorsque des protons sont observé intacts après interaction dans le détecteur AFP. Ce projet Forward Physique ATLAS permet l'extension du potentiel de découvert en physique à ATLAS, en permettant l'observation et la mesure d'une gamme de processus où un ou deux protons reste intact et qui serait difficile ou impossible d'étudier autrement. Ce chapitre se concentre sur la recherche de couplages de jauge quartic anormaux entre le photon et le W boson ($\gamma\gamma WW$).

L'étude de la désintégration semi-leptonique lors de la production de paires de WW par échange de photons : $pp \to p\gamma\gamma p \to pW^+W^-p$, a été réalsiée avec le générateur FMPC et la reconstruction rapide d'ATLAS. Le résultat montre des performances similaires par rapport aux études de désintégration leptonique pour la mesure du couplage anormal dans ATLAS utilisant les détecteurs AFP. La combinaison avec le mode de désintégration dileptonique promet d'augmenter la résolution de la mesure du terme de couplage quartique anormal a_0^W.

Cependant, l'étude complète de ce processus dans ATLAS consistant à intégrer les événements de pile-up et la reconstruction complète reste à faire. D'après les résultats préliminaires présentés ici, nous pensons que l'utilisation d'une approche similaire à l'œuvre présentée pour le cas leptonique - en utilisant 10 picosecondes de résolution pour les détecteur de timing afin d'identifier les vertex d'interaction primaire ainsi que la reconstruction des dijets à leur vertex d'origine - permettrait l'obtention d'une résolution comparable à celle présentée ici.

10.3 Photo-détecteurs rapide et algorithmes pour mesure du temps à la picoseconde

10.3.1 Photo-detecteurs à galette de micro-canaux

Dans ce chapitre, nous décrivons un photo-détecteur prometteur pour la mesure de temps de vol picoseconde: le photo-détecteur à galette de micro-canaux. Ce détecteur convertit rapidement un photon en un amas de 10^5 à 10^7 électrons, mesurable par une électronique dédiée externe. Comme nous le verrons, le procédé rapide garantit une faible variabilité de mesure du temps entre le moment de l'incidence des photons sur le détecteur et la mesure du signal électrique, ce qui garantie une mesure précise du temps (de l'ordre de la picoseconde).

Ce chapitre est divisé en huit sections. La section 5.2 présente les bases des photo-détecteurs, la section 5.3 la structure des détecteurs à galette de micro-canaux, et ses différents composant: la fenêtre d'entrée section 5.4, la section 5.5 présente la photocathode, les galette de micro-canaux sont présentée 5.6. Enfin, les sections 5.7 et 5.8 présentent les propriétés temporelle des détecteurs à galette de micro-canaux et leurs système de lecture par réseau d'anodes section 5.9.

Le photo-détecteur à galette de micro-canaux qui est présenté dans cette thèse est un détecteur de verre carré de 20cm de coté et à été mis au point par la collaboration Large Area Picoseconde Photo-détecteur (LAPPD). Cette collaboration se compose de physiciens et d'ingénieurs répartis dans trois laboratoires nationaux des États-Unis; Argonne, Fermilab et SLAC, cinq universités : UC Berkeley, Chicago, Illinois à Chicago, Illinois Champaign-Urbana, Hawaii et Washington à Saint-Louis ainsi que trois petites entreprises aux États-Unis : Arradiance, muons Inc et Synkera.

La première année de cette thèse a été passé à l'Université de Chicago en tant que membre de la collaboration LAPPD, où l'anode du détecteur (voir la section 5.9) a été étudié et sa bande passante améliorée. Une contribution importante a été apporté à l'électronique de lecture envisagée pour équipper le détecteur : la famille des ASIC PSEC (voir le chapitre 8).

La conception photo-détecteur à galette de micro-canaux LAPPD est fortement motivée par le développement de photo-détecteurs grande surface avec une excellente résolution temporelle (quelques picosecondes). Le projet Forward Physique ATLAS (voir le chapitre 3) bénéficierait grandement des avancées aportées par la collaboration LAPPD. En effet, ce type de détecteur pourrait être utilisé en association avec les radiateurs de quartz dans le détecteur de temps de vol.

Le projet de LAPPD, dont je fus membre pendant deux ans, et en particulier au cours de la première année de ce travail de thèse, est toujours un projet en cours.

La plupart des choix de conception furent des conpromis, entre le développement photo-détecteurs robustes, de grande surface et bon marché, (par exemple en utilisant un substrat de verre peu onéreux: le Borofloat33 de Schott, ou bien un système de lecture grande surface via des anodes microstrip) et essayer d'obtenir la meilleure résolution temporelle (recherche de la meilleure géométrie des pores dans les galettes de micro-canaux ou en essayant d'obtenir la bande passante la plus importante pour les anodes).

Aujourd'hui, ce projet est une réussite, avec un premier join étanche pour une surface de 20 cm par 20 cm et un photo-détecteur complètement fonctionnel. Ainsi, cela ouvre de nouvelles perspectives d'instrumentation de grandes surface de détection en physique des particules. Les démonstrations de résolution temporelle et les excellentes résolutions spatiale obtenues nous permettent d'envisager le détecteur LAPPD comme le candidat idéal comme photo-détecteur temps-de-vol picoseconde.

10.3.2 Mesure du temps à la picoseconde

Dans ce chapitre, nous étudions le moyen d'extraire l'information temporelle d'un signal étudié. Tout d'abord, nous décrivons ce qu'est le temps d'un événement et la façon de le référencer. Ensuite, nous présentons deux algorithmes utilisés pour extraire les informations temporelle d'un signal et discutonsr de leur précision relative. En fin de compte, nous allons présenter les paramètres clés de signaux arbitraires et leur impact sur la résolution temoporelle ultime. Pour ce faire, une simulation MATLAB permettant d'extraire les informations temporelle est présentée.

Pour résumer, nous avons obtenu une idée claire des facteur slimitant pour extraire un moment précis sur un signal échantillonné. Ils sont exprimés simplement par l'écart type σ_t, ou la variance donnée par l'algorithme de corrélation croisée:

$$\boldsymbol{\sigma_t \sim \sqrt{\frac{nt_{step}^2\sigma_n^2}{4A^2} + \frac{\sigma_j^2}{n}}} \tag{10.1}$$

Ce résultat est en bon accord avec les résultats rapportés dans la littérature, la mesure de temps précise s'appuie sur:

- Bon rapport signal à bruit.
- Un faible jitter d'échantillonnage.
- Un grande bande passante analogique
- Des signaux rapides.

Ce résultat nous motive dans la recherche d'une grande bande passante analogique et d'un faible bruit dans la conception d'un ASIC pour l'échantillonnage des photo-détecteurs rapides.

10.4 Dessin d'ASIC de mesure de temps picoseconde et résultats

Dans le chapitre **??**, nous présentons la puce SamPic, qui a été conçu, fabriquée et testé au CEA de Saclay. L'objectif de cette puce est de réaliser une mesure du temps en utilisant les techniques d'estimation présentée dans le chapitre 6. Le but ici est donc de concevoir un dispositif capable d'échantillonner très rapidement par rapport a une référence de temps précise, et d'enregistrer à la fois les échantillons et le temps

de référence. Le concept a été introduit dans le chapitre 6, mais est montré ici à nouveau dans la Figure 10.1.

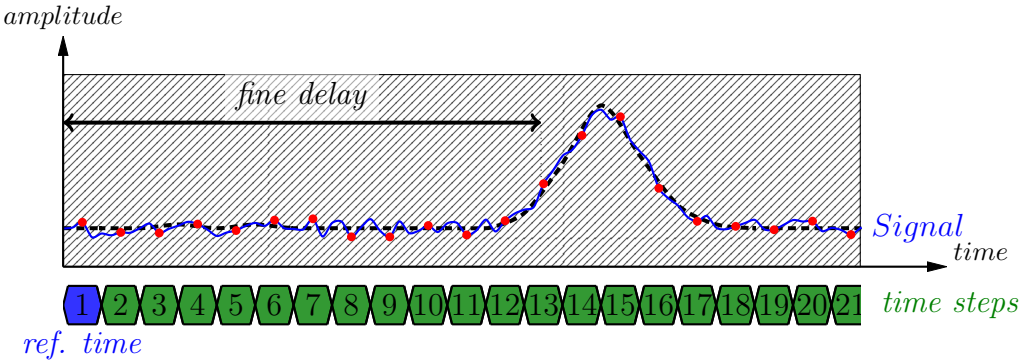


Figure 10.1 – L'objectif de l'ASIC est d'échantillonner (points rouges) le signal continu (en bleu) a des instants précis (en vert). L'échantillonnage est effectué par rapport à une horloge parfaitement connue, offrant une excellente résolution temporelle par rapport au premier point d'échantillonné *ref. time* (en bleu). À la fois les échantillons et le temps de référence sont enregistrées.

La conception est faite avec un soin particulier afin de minimiser le bruit d'échantillonnage et le jitter d'échantillonnage, qui sont les deux principaux paramètres pour la mesure précise du temps (voir le chapitre 6).

La conception de la puce SamPic a été lancée au cours de la deuxième année de cette thèse. Ele tient compte des leçons apprises au cours de ma contribution aux puces de la famille PSEC dévelopée à l'Université de Chicago La puce PSEC-4, dernière de la famille, a obtenu de très bons résultats, à la fois dans la vitesse et le bruit d'échantillonnage. Ces résultats ne sont pas présenté ici, mais ses caractéristiques peuvent être trouvés dans la littérature.

La conception de SamPic n'est pas une simple évolution de la puce PSEC-4, qui a été fabrqué dans le processus CMOS $0,13\mu m$ d'IBM. Des options de conception différentes ont été prises pour la rendre plus robuste et afin d'obtenir de meilleures performances. L'une des principales différences est également le choix d'une technologie avec une plus grande largeur de grille, offrant des fuites plus faible, une plage dynamique supérieure à un prix quatre fois plus faible.

De plus, la puce SamPic a été conçu comme un système de mesure de temps complet, directement utilisable dans des expériences ayant de fort taux de comptage.

Les puces SamPic et ASIC PSEC ont été développés pour une mesure du temps précise, mais tout particulièrement pour enregistrer des signaux provenant du photodétecteur LAPPD (chapitre 5) ou le détecteur de timing de la station AFP (chapitre 3). Par conséquent, une attention particulière a été accordée aux caractéristiques du détecteur durant toutes les étapes de la conception, afin de développer une électronique compatibles et directement intégrables dans les expériences.

10.4.1 SamPic

Comme décrit dans le chapitre 7, nous présentons ici les travaux sur la conception de l'ASIC SamPic. SamPic signifie Sampler for picoseconde time pick-off. Comme son nom l'indique, l'objectif de cet ASIC est de mesurer précisément le temps, c'est à dire échantillonner des signaux ultra-rapides avec un faible bruit et un faible jitter, référencés à une base de temps absolue très précise.

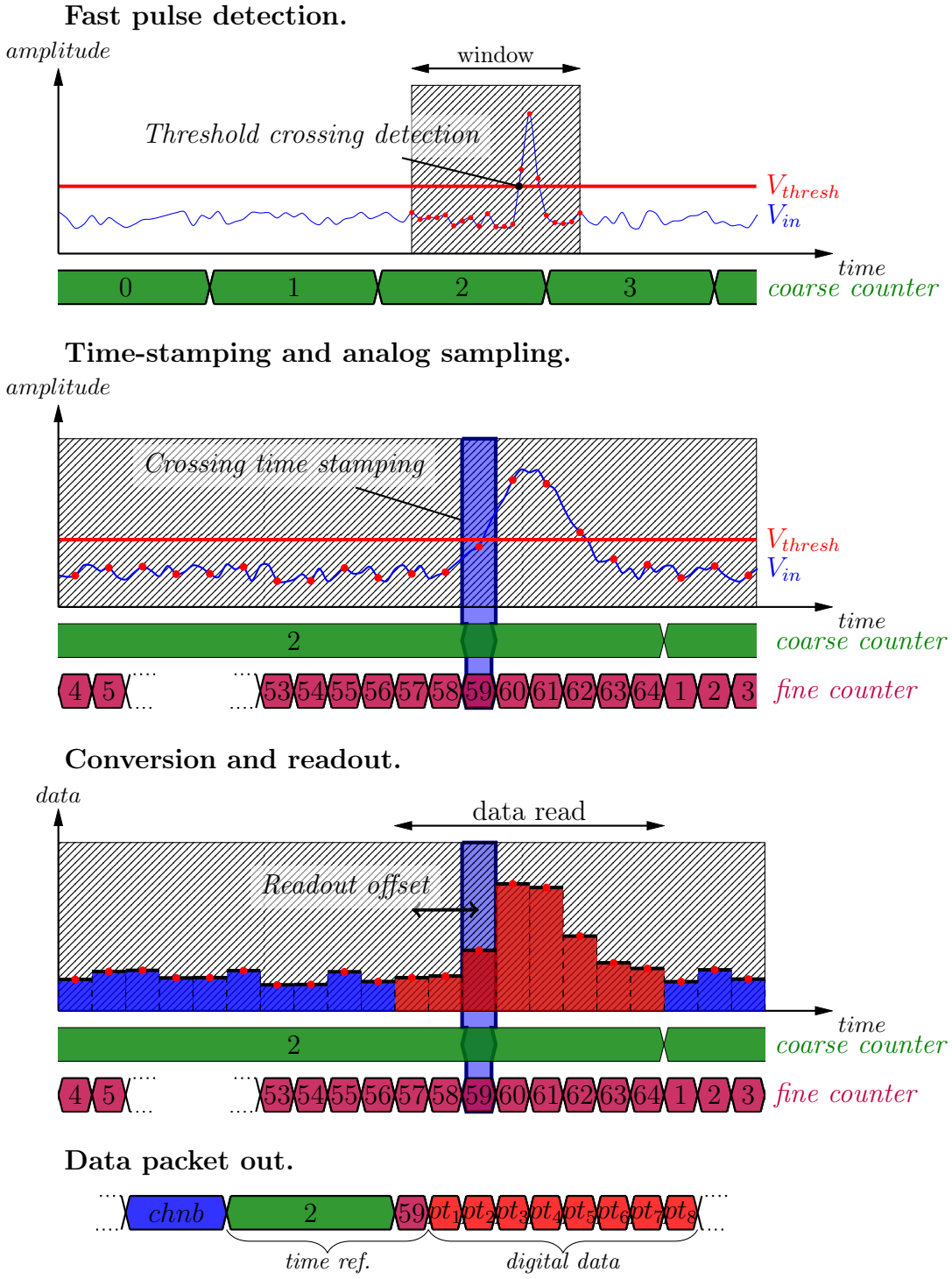


Figure 10.2 – Le principe de SamPic est illustré ici: détection de pulses rapide, datage temporel et l'échantillonnage analogique, la conversion et la lecture

La puce SamPic détecte, date, échantillonne et convertit des impulsions analogiques rapides. Le principe de son fonctionnement est illustré par la Figure 10.2.

La puce SamPic peut traiter 16 signaux d'entrée différents dans autant de canaux, (section 8.4). L'architecture générale de la puce est montrée Figure 10.3.

Références temporelles À l'intérieur de la puce, les instants d'échantillonnage sont définis par une *ligne à retard*, verrouillée à une horloge de référence externe Ck_{ref} (section 8.7).

Deux bases de temps sont utilisés pour garder la trace du temps écoulé :

- Une base grossière *compteur Gray* qui échantillonne de manière précise l'horloge de référence Ck_{ref}.
- Une base de temps fine avec la granularité de la ligne à retard.

Canaux Dans chaque canal, le franchissement du seuil est détecté par le *discriminateur interne* (section 8.10).

Chaque canal mémorise en permanence des échantillons analogiques du signal d'entrée dans un buffer FIFO de *cellules d'échantillonnage analogique* (section 8.6).

Lorsqu'un trigger se produit, les cycles d'écriture dans le répertoire *cellules d'échantillonnage analogique* sont arrêtés et un *hit* est enregistré (section 8.10).

- Les valeurs de la *position de la DLL* et du *compteur grossier* sont latchés.
- Les valeurs des *cellules d'échantillonnage analogique* sont numérisés par l'*ADC interne* (section 8.9).

Le déclenchement, la conversion et la lecture des cellules pour chaque canal est géré par la *contrôleur de canal*.

Lecture de la puce Le *contrôleur de la puce* fournit le trigger, la conversion et le signal de lecture à chaque canal (section 8.12).

SamPic est fabriqué avec le kit de conception CMOS haute tension $0.18\mu m$ fourni par l'Austria Micro System - cependant, la partie haute tension du kit n'est pas utilisé. Les détails des caractéristiques et des performances de cette technologie sont confidentielles.

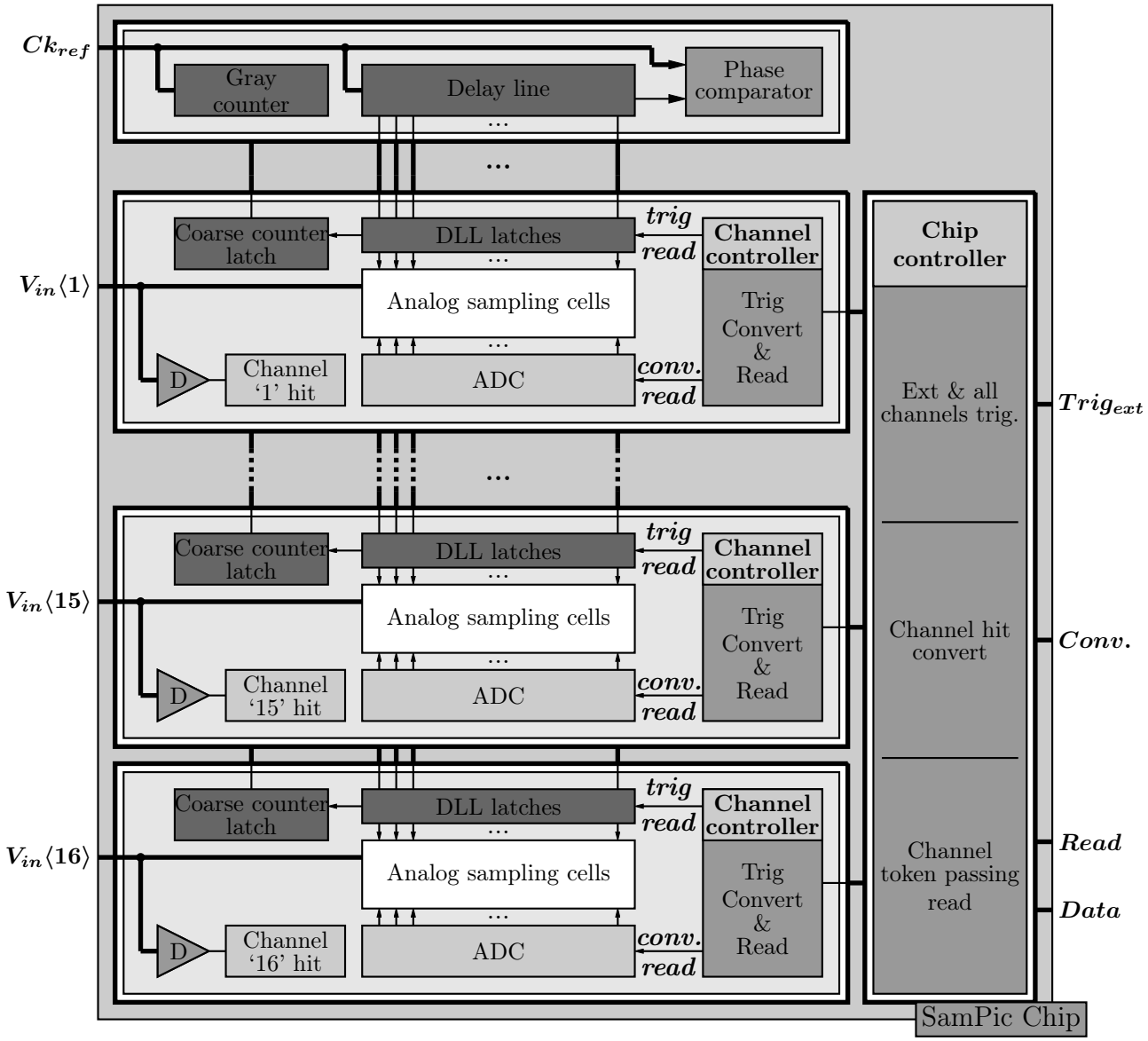


Figure 10.3 – Schéma de haut niveau de la puce SamPic.

Il y a un discriminateur par canal. Le rôle du discriminateur est de détecter un pulse rapide à l'entrée du canal. Il permet de définir une fenêtre autour du signal d'intérêt (par franchissement de seuil) où 64 échantillons du signal seront enregistrés comme le montre la Figure 10.4). Les détails sont décrits dans la section 8.10.

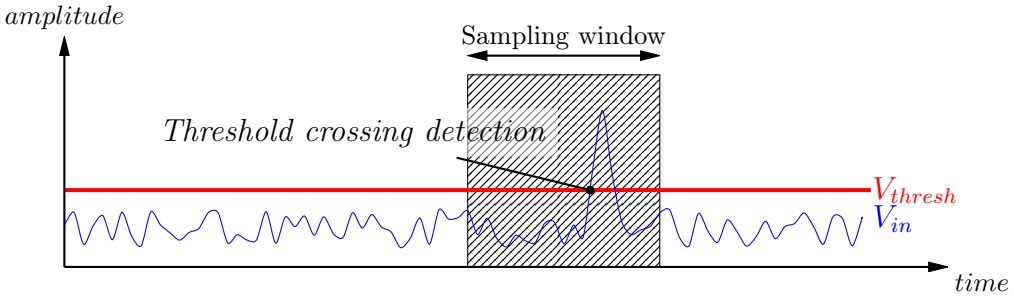


Figure 10.4 – Le rôle du discriminateur interne à chacune des voies de SamPic est de détecter des impulsions rapides en entrée.

La mise en mémoire tampon à l'intérieur de chaque canal analogique de SamPic est réalisée par une matrice de 64 cellules d'échantillonnage. Le rôle de ces cellules est de séquentiellement suivre et d'échantillonner le signal d'entrée du canal et de stocker les 64 échantillons de données précédentes, comme montré Figure 10.5. Les détails sont décrits dans la section 8.6.

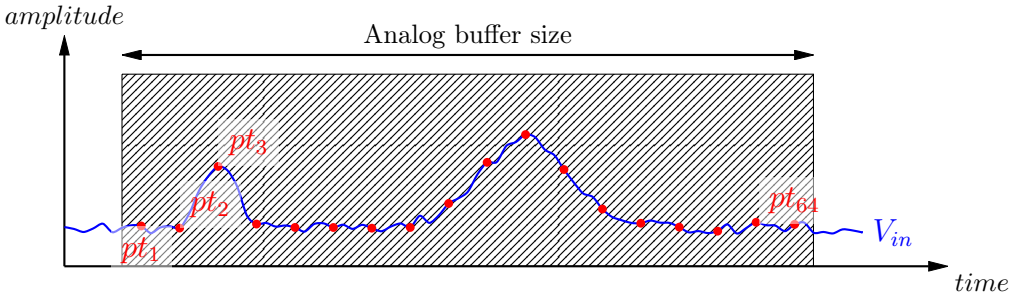


Figure 10.5 – Le rôle des cellules d'échantillonnage de SamPic est de bufferiser de façon continue des échantillons analogiques du signal d'entrée.

Le rôle des horloge lente et rapide à l'intérieur SamPic est de fournir une référence de temps pour les événements échantillonnés dans chaque canal, comme le montre la Figure 10.6. La base de temps rapide est une référence pour chaque point échantillonné dans SamPic tandis que la base de temps lente, fournie par l'horloge lente de SamPic, fournit une référence pour chaque «frame» (groupe de 64 échantillons) échantillonné dans SamPic.

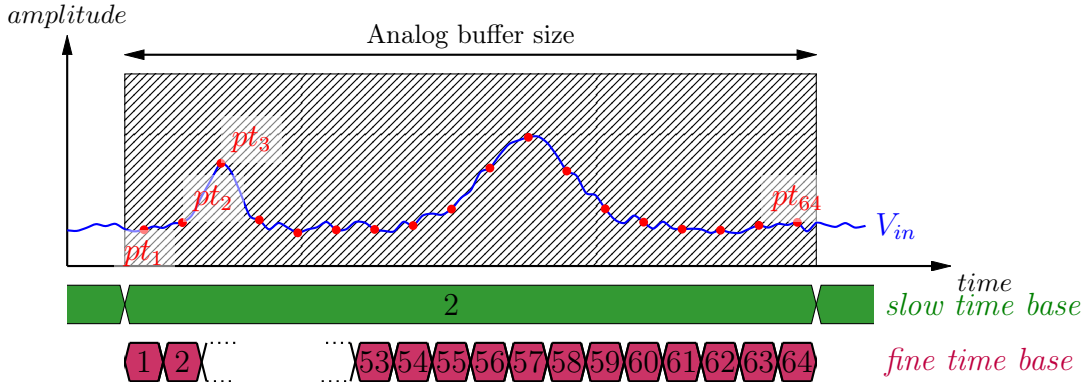


Figure 10.6 – La base de temps lente et rapide en SamPic fournir référence de temps précise pour tous les points échantillonnés.

Une ligne à retard asservie de 64 éléments (delay locked loop - DLL), dont le retard total est asservi à la durée de la base de temps lente est utilisée pour fournir les 64 signaux utilisés pour commander l'échantillonnage dans SamPic. Ces mêmes signaux sont utilisés pour déterminer la base de temps rapide. Le retard asservi garanti que les instants d'échantillonnage sont précis et régulièrement espacés, comme le montre la Figure 10.7. Les détails sont décrits dans la section 8.7.

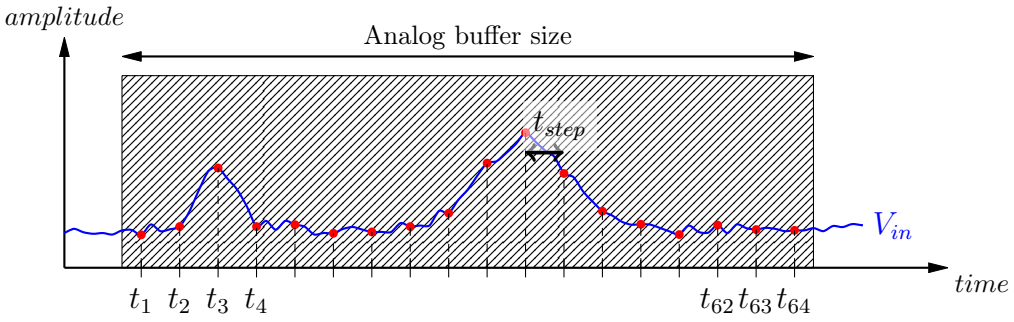


Figure 10.7 – Le rôle de la DLL de SamPic est de générer les instants d'échantillonnage rapides et précis.

Une fois échantillonnée, les valeurs analogiques sont convertis avec une résolution de 11 bits. C'est le rôle du convertisseur analogique-numérique (ADC) présent dans chaque canal, comme le montre la Figure 10.8. Les détails sont décrits dans la section 8.9.

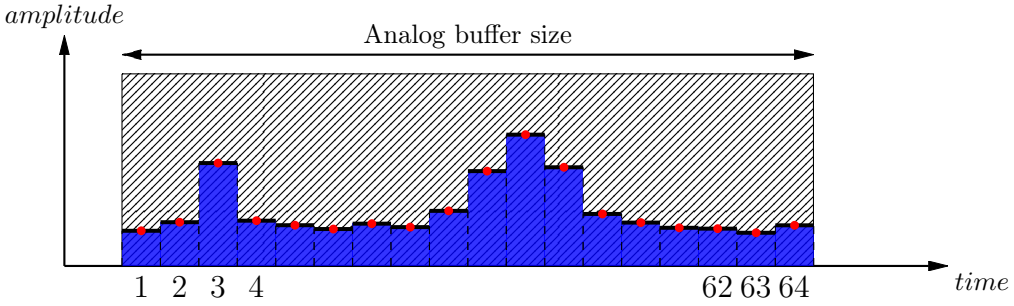


Figure 10.8 – Le rôle de l' ADC de chacune des chaînes de la SamPic est de convertir la valeur échantillonnée analogique en valeurs numériques.

Une fois converti, les données des 16 canaux convertis sont lus grace au systèm d'acquisition. Dans chaque canal le détecteur de région d'intérêt (Region Of Interest - ROI) permet la lecture intelligente des données d'intérêt (autour de l'évènement de trigger) et donc accélérer le processus de lecture, comme le montre la Figure 10.9.

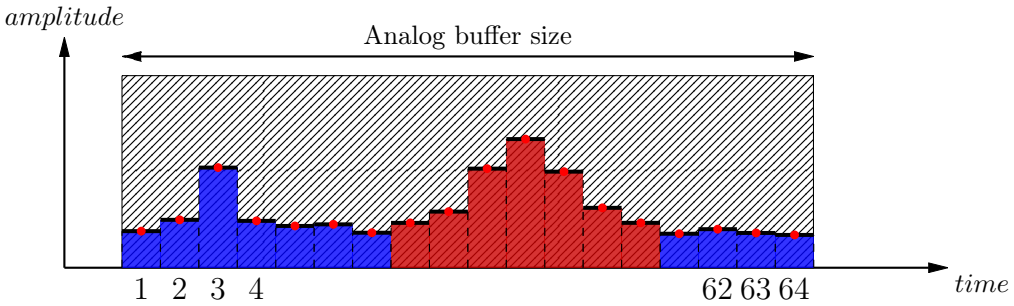


Figure 10.9 – Le rôle de la lecture par Région d'intérêt est de lire uniquement les données intérressante et donc accélérer le processus de lecture.

L'échantillonnage dans chaque canal est réalisé par un réseau de 64 cellules mémoires identiques abritant les éléments suivante : une cellule d'échantillonnage analogique (section 8.6), un convertisseur analogique-numérique Wilkinson (section 8.9), le stockage de données numériques sur 11 bits (section 8.9) et un générateur de trigger pour la cellule d'échantillonnage. Le schéma de principe des canaux est montré dans la Figure 10.10. Les détails sont décrits dans la section **??**.

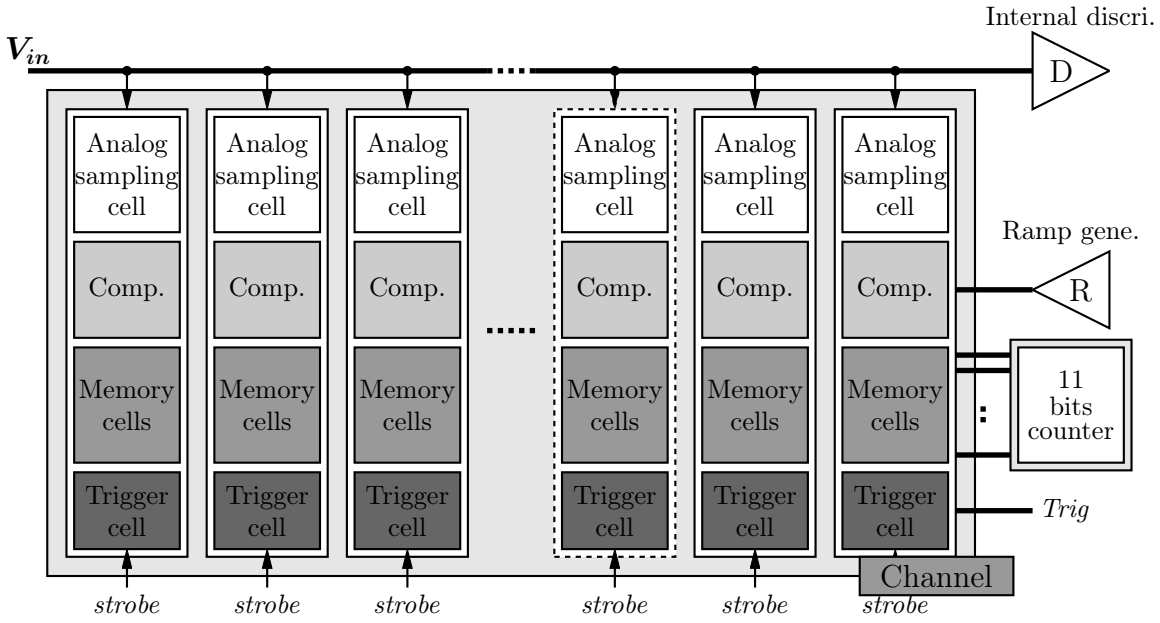


Figure 10.10 – Schéma d'un canal SamPic. Chacun se compose d'un groupe de 64 cellules mémoire, chacune comporte une cellule d'échantillonnage analogique, un convertisseur Wilkinson, une cellule de stockage numérique et une cellule de trigger.

L'échantillonnage dans chaque canal est réalisée par une matrice de 64 cellules d'échantillonnage, ce qui est une valeur assez faible. Cette décision a été prise comme une conclusion de l'atelier Timing à Chicago : un petit nombre de cellules d'échantillonnage augmentant la compacité du canal, et donc la bande passante. Chaque cellule est commandée individuellement par une *cellule de trigger* (voir la section 8.10) qui utilise les signaux de la ligne à retard (voir la section 8.7) et le signal de trigger. L'architecture est représentée sur la Figure 10.11.

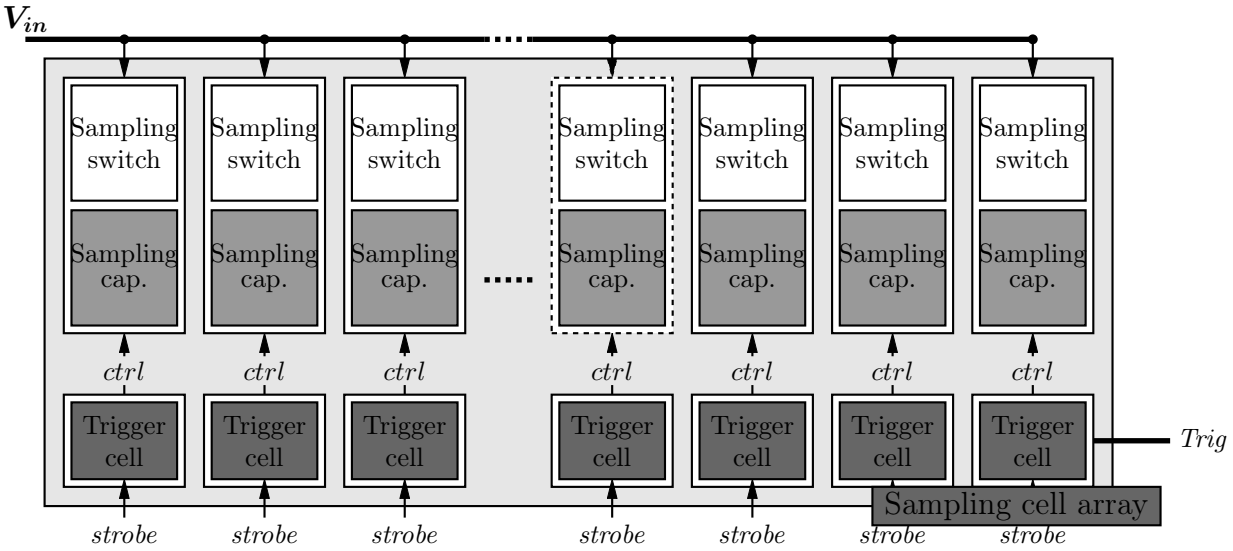


Figure 10.11 – Schéma des cellules d'échantillonnage dans le canal SamPic. Il ya 64 cellules d'échantillonnage par canal. Chaque cellule d'échantillonnage est constitué d'un switch d'échantillonnage et d'une capacitér d'échantillonnage.

Le principe de l'échantillonnage à l'intérieur d'une cellule d'échantillonnage est décrit dans la Figure 10.12. Le signal de commande de la cellule de trigger, met la cellule en mode *suiveur* quand il est haut. Dans le mode suiveur la tension dans la cellule suit la tension d'entrée V_{in}. Lorsque la tension de contrôle passe au niveau bas, la cellule d'échantillonnage passe du mode suiveur au mode échantillonné. Le principe est illustré dans la Figure 10.12.

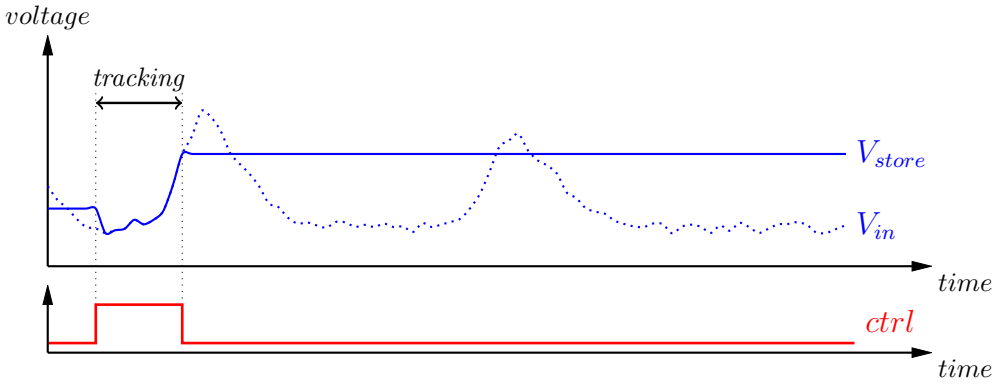


Figure 10.12 – Principe de l'échantillonnage dans la cellule d'échantillonnage. Dans le mode suiveur de la tension dans la cellule suit la tension d'entrée V_{in}. Lorsque tension de contrôle passe au niveau bas, la cellule d'échantillonnage arrête de suivi de la tension d'entrée et garde sa dernière valeur.

Le principe de l'échantillonnage dans de multiples cellules connectés à une même ligne d'entrée est illustré dans la Figure 10.13. Ici, quatre cellules sont connectées et contrôlées par quatre signaux de contrôle identiques décalés dans le temps par t_{step}. Cela permet d'échantillonner quatre points uniformément espacés (pt_1 à pt_4) sur le signal d'entrée V_{in}.

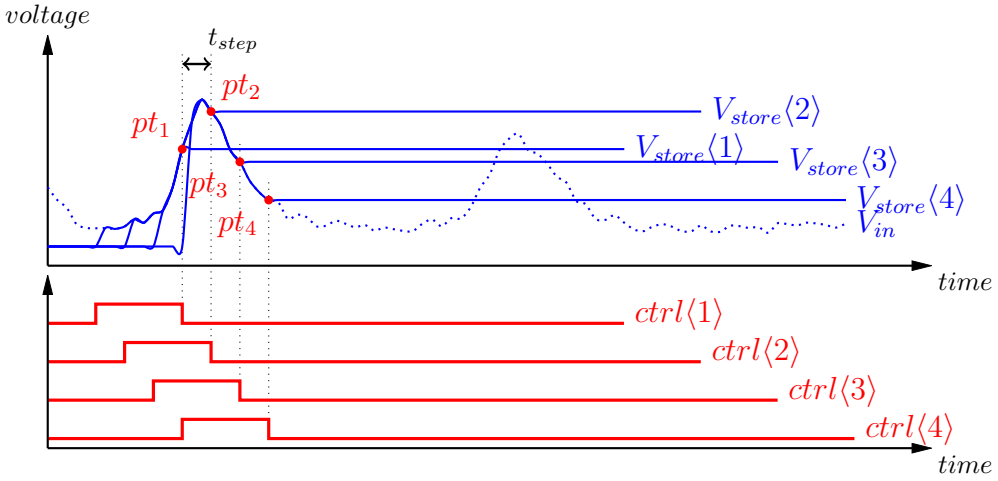


Figure 10.13 – Principe de l'échantillonnage avec un groupe de quatre cellules. Ici 4 cellules sont reliées et contrôlées par quatre signaux de commande identique décalée dans le temps par t_{step}.

Afin de bufferiser le signal analogique d' entrée V_{in} de manière continue, les cellules d'échantillonnage sont réécrites à partir du moment où toutes les cellules ont été écrites, ce qui créé une mémoire tampon circulaire. Ceci est illustré dans la Figure 10.14, avec 6 cellules. Pour assurer la continuité du buffering dans le réseau de cellules, les commandes de contrôle des cellules d'échantillonnage sont répétées au fil du temps. Si nous avons nb_{cell} dans la mémoire tampon, le temps de répétition t_{rep} doit être :

$$t_{rep} = nb_{cell} \times t_{step} \tag{10.2}$$

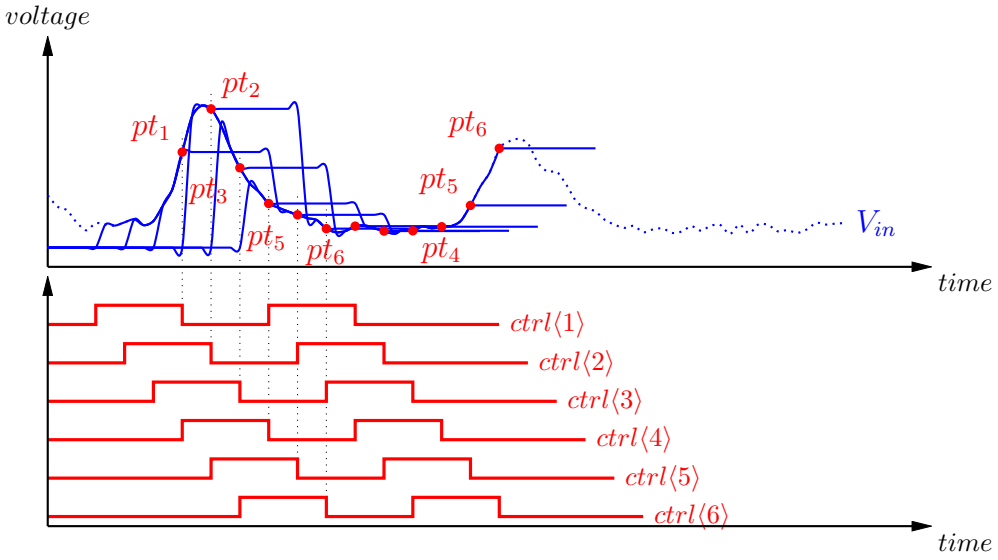


Figure 10.14 – Principe de mise en mémoire tampon analogique avec un réseau de cellules d'échantillonnage analogiques. Afin d'enregistrer en continu le signal analogique d'entrée V_{in}, les cellules d'échantillonnage sont réécrites au moment où toutes les cellules sont pleines, ce qui rend la mémoire tampon circulaire.

À l'intérieur de la mémoire des cellules sont en mode suiveur pendant $t_{tracking}$, tandis que les autres garde leur valeur échantillonnée.

Les cellules en mode suiveur dans la mémoire ne détiennent pas de données précieuses, car ils sont tous reliés à la ligne d'entrée et partagent la valeur $V_{in}(t)$. Le reste des cellules $nb_{cell} - nb_{track}$ sont dans le mode échantilloné et gardent la mémoire de la valeur précédente de V_{in}. Afin de bufferiser efficacement le signal d'entrée, le nombre de cellules dans le mode suiveur nb_{track} doit être une faible fraction du nombre total de cellules nb_{cell}.

Ce principe est illustré dans la Figure 10.15. Dans cet exemple, 5 cellules mémoire suivent simultanément le signal d'entrée, tandis que les cellules $64 - 5 = 59$ sont mémorisent les valeurs d'entrée précédentes. Si l'échantillonnage arrêté brutalement, ces 5 cellules mémoire passeront simultanément du mode suiveur au mode échantillonné, avec toutes la même valeur, c'est ce que nous appelons: cellules mortes.

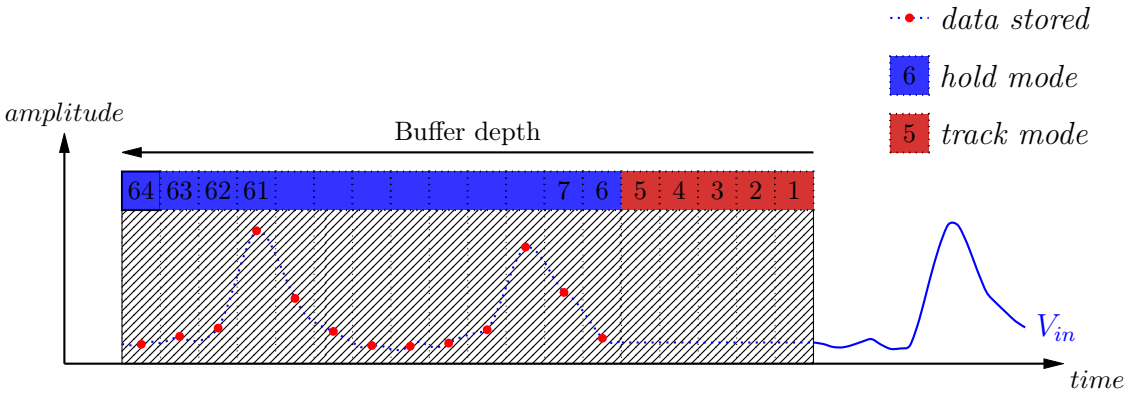


Figure 10.15 – Principe de mémoire analogique avec un réseau de cellules d'échantillonnage.

Nous pouvons voir que les cellules mortes sont une limitation des mémoires analogique, en particulier dans le cas de mémoire de faible profondeur.

Dans la section 8.10, une méthode est présentée qui supprime entièrement les cellules mortes. Cependant, les derniers échantillons sont prélevés après le signal de déclenchement seulement.

Une séquence d'échantillonnage est présentée à titre d'illustration dans la Figure 10.16 pour temps d'échantillonnage consécutifs. Dans cette illustration, 5 cellules suivent constamment le signal d'entrée, à savoir: $t_{tracking} = 5 * t_{step}$ et apparaissent comme des cellules mortes dans la mémoire analogique.

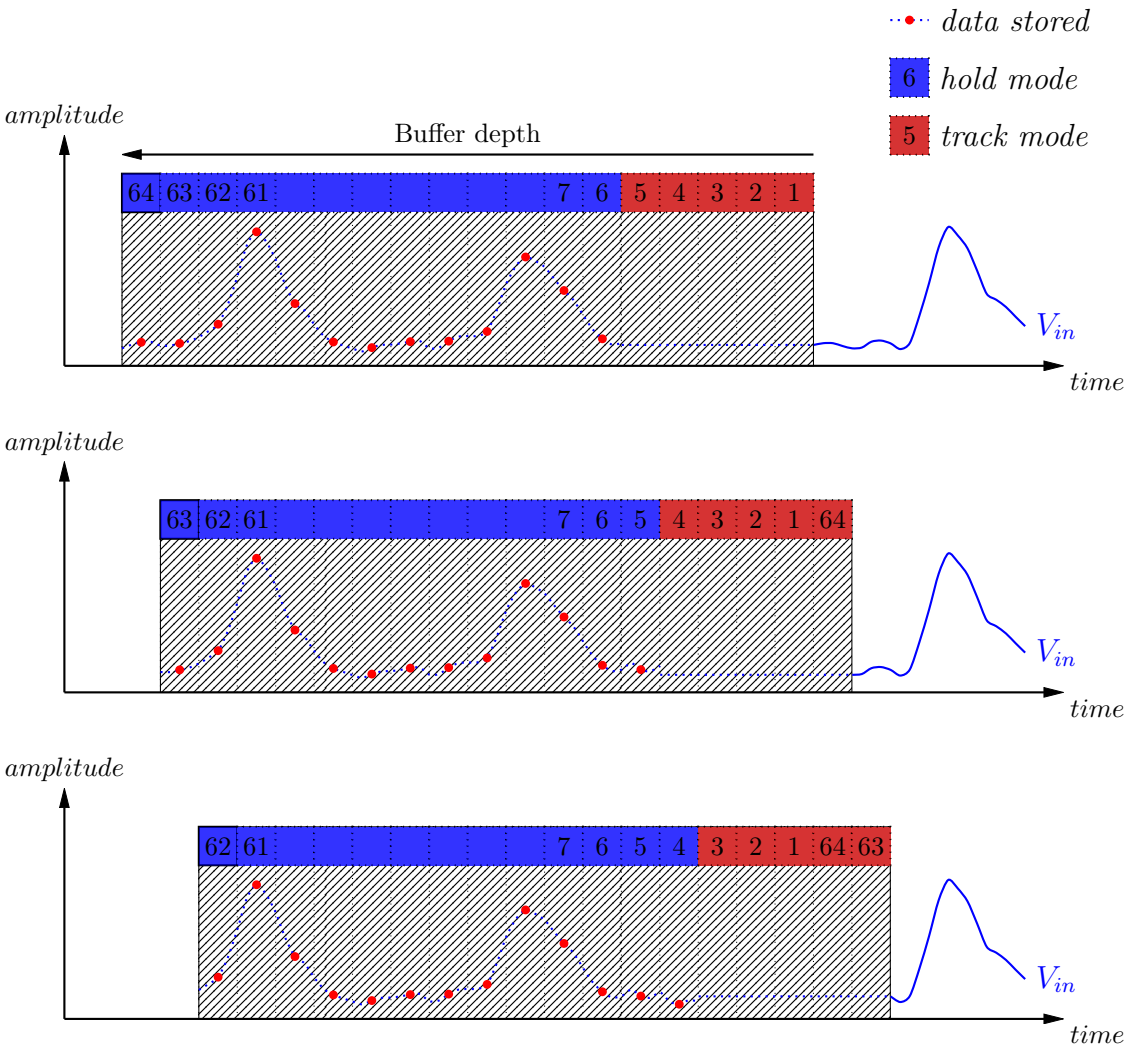


Figure 10.16 – Illustration de la séquence de mise en mémoire tampon pour 3 temps d'échantillonnage successif t_{step} dans une mémoire analogique de 64 cellules avec 5 cellules connectées au bus d'entrée.

Dans SamPic nous avons choisi une mémoire constituée de 64 cellules d'échantillonnage analogique. Ce nombre est choisi pour être une puissance de deux, afin de simplifier l'adressage et le codage dans la puce qui utilise le code binaire. Le nombre de 64 nous donne une profondeur de mémoire suffisante pour enregistrer les données d'intérêt d'un signal rapide. Ce qui permet de respecter les exigences de taille, puissance et temps de lecture de la puce.

Une illustration de SamPic profondeur d'échantillonnage est illustrée à la Figure **??**. A l'intérieur des canaux de SamPic nous pouvons mémoriser 64 points sur le signal d'entrée. Tous les points sur des données antérieures seront écrasés et perdus.

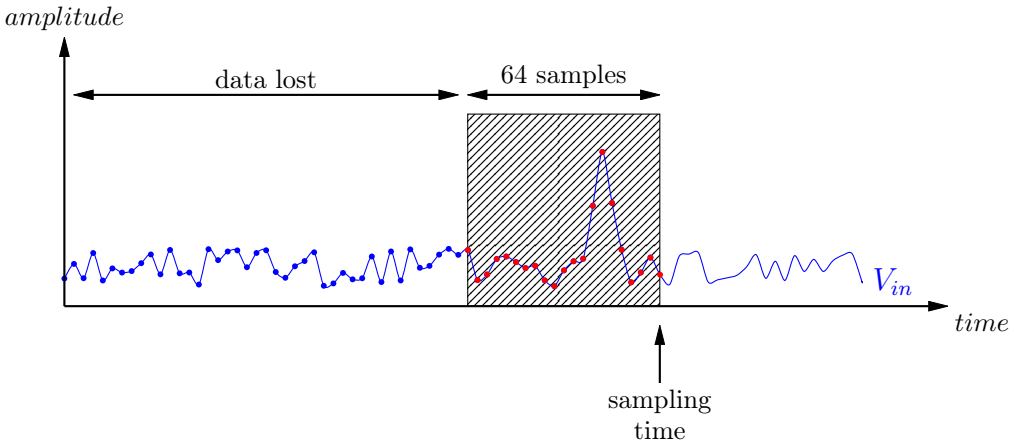


Figure 10.17 – Illustration du principe d'échantillonnage SamPic et la profondeur de la mémoire tampon de 64 points. 64 points (rouges) sont continuellement échantillonné sur le signal V_{in} à chaque instant d'échantillonnage. Les points les plus anciens (bleus) sont écrasés et perdus au fur et à mesure de l'acroissement des instants d'échantillonage.

SamPic est une puce directement destinée à mesurer le temps d'un signal. Au cours de toutes les étapes de la conception une attention particulière a été accordée à tous les paramètres et les caractéristiques incidence sur la sensibilité de la mesure du temps tels que le bruit, le jitter ou les décalages. La puce a été soumis avec succès à la fonderie et de fabrication. Une image de l'agencement d'une puce complète peut être vu dans la Figure 10.18

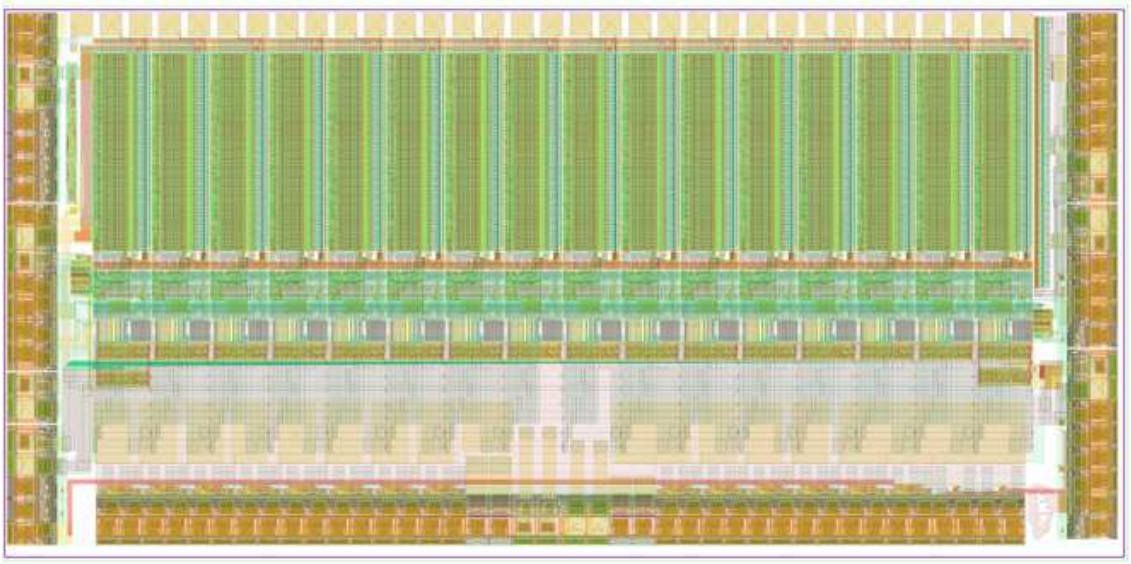


Figure 10.18 – Image du layout de la puce SamPic.

10.5 Résultats

Les résultats préliminaires présentés ici sont très encourageants en particulier la précision atteinte sur les mesures de temps. Un tableau résume les principales performances de la puce SamPic.

Table 10.1 – Table des principales performance de SamPic.

	Performances mesurée
Nombre de canaux	16
Bande passante 3dB	1.6 GHz
Fréquence d'échantillonnage	3.2 - 10 GSa.s^{-1} [2]
Nombre de points	64
ADC precision	11 bits
Bruit	1 mV RMS
Dynamique	1 V
Horloge de lecture	160 MHz [3]
Temps de conversion	1.6 μs pour 11 bits - 200 ns pour 8bits
Vitesse de lecture	1.92 Gb.s^{-1}

10.6 Conclusion

Dans cette thèse de doctorat, nous avons présenté la conception d'un système de mesure temps de vol picoseconde pour l'expérience ATLAS. Ce travail de recherche a été fondée à la fois sur les projects Large Area Picosecond Photo-Detector, à l'Université de Chicago, et SamPic, au CEA.

L'étude de la sensibilité à de la physique au-delà du Modèle Standard, apporté par la conception et l'installation de détecteurs picoseconde temps de vol dans la région avant de l'expérience ATLAS au LHC et présentée dans la première partie de la thèse montre que l'utilisation des détecteurs AFP au LHC apportera une amélioration significative de la mesure de couplages de jauges quartic anormaux entre le photon et le boson W, cela par l'observation de la production exclusive de paire WW dans ATLAS. La désintégration semi-leptonique considéré pour la paire WW, permet de sonder des couplages anormaux a_0^W de l'ordre de 5.10^{-6}, similaire au cas leptonique pur, et apportant ainsi une sensibilité supplémentaire dans la recherche de couplages quartic anormaux au LHC .

Afin de pouvoir réaliser cette mesure, un détecteur de trace et de timing précis est nécessaire dans ATLAS : le détecteur AFP. Le déploiement du détecteur AFP dans le tunnel du LHC est présenté dans la première partie de la thèse. Dans la deuxième et troisième partie, nous présentons une description des photo-détecteurs picoseconde grande surface, les algorithmes de reconstruction de temps et les circuits intégrés de

[2] Seuls les 8 premiers canaux sont fonctionnel à 10 GSa.s^{-1}

[3] 400Mhz possible non vérifié

lecture dédiés. Un soin tout particulier apporté à l'échantillonnage du signal et le traitement pour une mesure de temps précise.

La description des photo-détecteurs picoseconde grande surface est basée sur le travail de recherche et de développement réalisé à l'Université de Chicago, au sine de la collaboration LAPPD. Une attention particulière a été donnée à toutes les parties du détecteur critiques pour un timing précis. Aujourd'hui le détecteur 20 par 20 cm est très mature, en particulier, toutes les parties actives : photo-cathode, galette de micro-canaux, réseau d'anode et la lecture électronique conçu avec succès. des impulsions avec temps de montée inférieurs à 1 ns ont été observées avec une source laser pointée sur le détecteur et un vide appliqué de manière externe.

Afin de faire face au taux élevé et à faible temps mort nécessaire pour le détecteur AFP à l'expérience LHC un circuit intégré dédié (ASIC) appelé SamPic a été conçu pour fournir une lecture des données du détecteur de timing. L'ASIC a été conçu, fabriqué et testé avec succès. Il présente la meilleure précision dans ce domaine avec une résolution temporelle impressionnante entre 5 à 7 picosecondes pour signaux retardés de 0 et 64 ns .

Le domaine de la mesure de temps précise gagne en popularité non seulement dans le monde de la physique des particules, mais plus généralement dans le domaine de l'imagerie ; une précision de 10 picosecondes en temps de vol pour la lumière est équivalent à une résolution en profondeur de 3 mm. L'industrie est de plus en plus attirés par cette nouvelle technologie afin d'obtenir des imageurs médicaux ayant des résolutions toujours plus grande, possible pr exemple à l'aide de PET à temps de vol. Ces améliorations pourrai aussi donner le jour à des télémètres plus précis, des lunettes de vision nocture plsu performante, mais aussi permettre de l'imagerie 3-D.

Hervé GRABAS

Développement d'un système de mesure de temps de vol picoseconde dans l'expérience ATLAS

Résumé

Cette thèse présente une étude de la sensitivité à la physique au delà du Modèle Standard au CERN grâce à la conception et l'installation de détecteur temps de vols picoseconde dans la région avant du détecteur ATLAS.
La première partie présente une étude, basée sur la production exclusive de paire WW au LHC, à la sensitivité aux couplages de gauge quartic entre le photon et le boson W. La sélection du signal considère la désintégration semi-leptonique de la paire WW et la présence du détecteur AFP dans ATLAS.
La deuxième partie présente la conception de SamPic: un circuit intégré spécifique pour la mesure picoseconde. Enfin, ses premiers résultats sont reportés, en particulier une précision exceptionnelle de 5ps sur la mesure de retard entre deux pulses.

Mots clés : Picoseconde ; mesure du temps ; temps-de-vol; photo-détecteur ; LHC ; couplages de jauge anormaux.

Abstract

In this thesis, we present a study of the sensitivity to Beyond Standard Model physics brought by the design and installation of picosecond time-of-flight detectors in the forward region of the ATLAS experiment at the LHC.
The first part of the thesis present a study of the sensitivity to the quartic gauge anomalous coupling between the photon and the W boson, using exclusive WW pair production in ATLAS. The event selection is built considering the semi-leptonic decay of WW pair and the presence of the AFP detector in ATLAS.
The second part gives a description of large area picosecond photo-detectors design and time reconstruction algorithms with a special care given to signal sampling and processing for precision timing.
The third part presents the design of SamPic: a custom picosecond readout integrated circuit. At the end, its first results are reported, and in particular a world-class 5ps timing precision in measuring the delay between two fast pulses.

Keywords: Picosecond, timing, time-of-flight, photo-detector, LHC, Anomalous gauge couplings.